

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:30:20 ; Search time 1849 Seconds

(without alignments)
6847.244 Million cell updates/sec

Title: US-09-646-679-14

Sequence: 1 ggcacgagcgaatcgaga.....aaataagggttcctctacc 605

Scoring table:

IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:*

1: gb.pa:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg.hum:*
31: em.htg.in:*
32: em.htg.other:*
33: em.htgo.in:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	605	100.0	605	6	AX016311	AX016311 Sequence
2	605	100.0	622	8	AF057027	Brassica
3	297.4	49.2	818	6	AX016324	AX016324 Sequence
4	289.4	47.8	576	6	AX016332	AX016332 Sequence
5	171.2	28.3	110804	8	ATAC016829	Arabidops
6	106.2	17.6	1716	6	AX016320	AX016320 Sequence
7	52.2	8.9	7218	6	166494	Sequence 14
8	51.4	8.6	3666	1	AF319618	AF319618 Vbri fo
9	49.4	8.5	111752	9	AL596454	Human DNA
10	49.2	8.2	15342	3	AE001420	Plasmodu
11	49.2	8.1	77835	2	PFMAL19P2_3	Continuation (4 of
12	49	8.1	193488	30	AC018519	AC018519 Homo sapi
13	49	8.1	198509	2	AC018512	AC018512 Homo sapi
14	48.8	8.1	1324	6	AX016336	AX016336 Sequence
15	48.8	8.1	151498	9	AC099331	AC099331 Homo sapi
16	48.8	8.1	190748	9	AC010969	AC010969 Homo sapi
17	48.2	8.0	107770	2	AC098572	AC098572 Oryza sat
18	48	7.9	122583	9	AC092435	AC092435 Homo sapi
19	48	7.9	141710	2	AC012348	AC012348 Homo sapi
20	47.8	7.9	81202	2	AC068000	AC068000 Homo sapi
21	47.8	7.9	81202	2	AC068000	AC068000 Homo sapi
22	47.8	7.9	124057	9	AP000770	AP000770 Homo sapi
23	47.8	7.9	132441	2	AC027776	AC027776 Homo sapi
24	47.8	7.9	169494	9	AL450083	AL450083 Human DNA
25	47.8	7.9	175706	2	AC013798	AC013798 Homo sapi
26	47.8	7.9	177147	9	AC007539	AC007539 Homo sapi
27	47.8	7.9	179753	9	AC096734	AC096734 Homo sapi
28	47.8	7.9	180851	9	AL365179	AL365179 Human DNA
29	47.6	7.9	55149	9	AL606862	AL606862 Human DNA
30	47.4	7.8	226	8	AY018221	AY018221 Oryza sat
31	47.4	7.8	135997	2	AC015261	AC015261 Oryza sat
32	47.4	7.8	144645	9	AC026273	AC026273 Homo sapi
33	47.4	7.8	156387	2	AC013593	AC013593 Homo sapi
34	47.4	7.8	173733	9	AL583856	AL583856 Human DNA
35	47.4	7.8	166093	2	AC106013	AC106013 Homo sapi
36	47.4	7.8	173794	9	AC023108	AC023108 Homo sapi
37	47.2	7.8	15951	1	AE001073	AE001073 Archaeogl
38	47.2	7.8	161835	2	AC104794	AC104794 Homo sapi
39	47	7.8	97160	9	CNS06C7	AL389914 Human chr
40	47	7.8	149278	2	AC011602	AC011602 Homo sapi
41	47	7.8	183082	9	AC019183	AC019183 Homo sapi
42	46.8	7.7	141854	2	AC021595	AC021595 Homo sapi
43	46.8	7.7	161551	2	AC027509	AC027509 Homo sapi
44	46.6	7.7	159712	9	AP004219	AP004219 Homo sapi
45	46.6	7.7	177730	9	AC087844	AC087844 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AX016311
DEFINITION Sequence 14 from Patent WO9949046.
ACCESSION AX016311
VERSION AX016311.1 GI:10041874
KEYWORDS
SOURCE
ORGANISM

Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.

REFERENCE
1 (bases 1 to 605)

AUTHORS
TITLE
JOURNAL
PATENT: WO 9949046-A 14 30-SEP-1999;
ROBERTS JEREMY ALAN (GB); BIOGENMA UK LTD (GB); WYATT PAUL (GB);
WITTELAU CATHERINE (GB)

FEATURES
Location/Qualifiers

1..605
/organism="Brassica napus"
/db_xref="taxon:3708"
20..430

CDS

/note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC07684.1"
 /db_xref="GI:10041875"
 /translation="WATKSMGDEIKKKLUNLYVDDPLNLIIEKTIKAGISQT
 ANNEEAVIHRDGSSEFDLIMDKEMPERDGVSTYRKUREMEVSMIVGTSIADNE
 EERRAFMEAGINHLAPLTKDKIIPILINOLMDA"
 BASE COUNT 198 a 101 c 140 g 166 t
 ORIGIN

Query Match 100.0%; Score 605; DB 6; Length 605;
 Best Local Similarity 100.0%; Pred. No. 9.8e-138;
 Matches 605; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACGACGAGATCGAAGATGCGCAAAATCCATGCGAGATATCGAAAAATAAGAA 60
 Db 1 GGCACGACGAGATCGAAGATGCGCAAAATCCATGCGAGATATCGAAAAATAAGAA 60
 QY 61 GAACTAAACGCTGTGATCGTGAATGATGATCCCACTTAACCTTATATGAGAAGAT 120
 Db 61 GAACTAAACGCTGTGATCGTGAATGATGATCCCACTTAACCTTATATGAGAAGAT 120
 QY 121 CATCAAGCGATTGGGGGATTTTCACAGACGCAATTAAGGTGAGAGCGCAATATCAT 180
 Db 121 CATCAAGCGATTGGGGGATTTTCACAGACGCAATTAAGGTGAGAGCGCAATATCAT 180
 QY 181 CCACCGTACGCGGCTCATCTTTGACCTTATCTTATGATTAAGAAATGCCGAGAG 240
 Db 181 CCACCGTACGCGGCTCATCTTTGACCTTATCTTATGATTAAGAAATGCCGAGAG 240
 QY 241 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 300
 Db 241 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 300
 QY 301 GGTGACTTACCTGCTGACAAATGAAGAGAGGCGGCTTTCATGGAAGCTGACTTAA 360
 Db 301 GGTGACTTACCTGCTGACAAATGAAGAGAGGCGGCTTTCATGGAAGCTGACTTAA 360
 QY 361 CCATTCCTTGGCAAAACCGTTAACCAAGACAGATCATCTCTCATTAACCACTCAT 420
 Db 361 CCATTCCTTGGCAAAACCGTTAACCAAGACAGATCATCTCTCATTAACCACTCAT 420
 QY 421 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 480
 Db 421 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 480
 QY 481 TATGTATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 Db 481 TATGTATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 QY 541 GATATATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 Db 541 GATATATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 QY 601 CTACC 605
 Db 601 CTACC 605

RESULT 2
 AF057027 622 bp mRNA linear PLN 01-OCT-1998
 LOCUS Brassica napus response regulator protein mRNA, complete cds.
 DEFINITION AF057027
 ACCESSION AF057027
 VERSION AF057027.1 GI:3687687
 KEYWORDS
 SOURCE
 ORGANISM
 Brassica napus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica
 REFERENCE
 1 (bases 1 to 622)
 WhiteLaw,C.A., Paul,W., Jenkins,E.S., Taylor,V.M. and Roberts,J.A.

TITLE A mRNA encoding a response regulator protein from Brassica napus is
 up-regulated during pod development
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 622)
 AUTHORS WhiteLaw,C.A.
 TITLE Direct Submission
 JOURNAL Submitted (02-Apr-1998) School of Biological Sciences, University
 of Nottingham, Sutton Bonington Campus, Loughborough,
 Leicestershire LE12 5RD, England U.K.
 FEATURES
 source
 location/Qualifiers
 1..622
 /organism="Brassica napus"
 /db_xref="taxon:3708"
 CDS
 20..430
 /note="up-regulated during pod development"
 /codon_start=1
 /product="response regulator protein"
 /protein_id="AAC62225.1"
 /db_xref="GI:3687688"
 /translation="WATKSMGDEIKKKLUNLYVDDPLNLIIEKTIKAGISQT
 ANNEEAVIHRDGSSEFDLIMDKEMPERDGVSTYRKUREMEVSMIVGTSIADNE
 EERRAFMEAGINHLAPLTKDKIIPILINOLMDA"

BASE COUNT 215 a 101 c 140 g 166 t
 ORIGIN

Query Match 100.0%; Score 605; DB 8; Length 622;
 Best Local Similarity 100.0%; Pred. No. 9.8e-138;
 Matches 605; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGCACGACGAGATCGAAGATGCGCAAAATCCATGCGAGATATCGAAAAATAAGAA 60
 Db 1 GGCACGACGAGATCGAAGATGCGCAAAATCCATGCGAGATATCGAAAAATAAGAA 60
 QY 61 GAACTAAACGCTGTGATCGTGAATGATGATCCCACTTAACCTTATATGAGAAGAT 120
 Db 61 GAACTAAACGCTGTGATCGTGAATGATGATCCCACTTAACCTTATATGAGAAGAT 120
 QY 121 CATCAAGCGATTGGGGGATTTTCACAGACGCAATTAAGGTGAGAGCGCAATATCAT 180
 Db 121 CATCAAGCGATTGGGGGATTTTCACAGACGCAATTAAGGTGAGAGCGCAATATCAT 180
 QY 181 CCACCGTACGCGGCTCATCTTTGACCTTATCTTATGATTAAGAAATGCCGAGAG 240
 Db 181 CCACCGTACGCGGCTCATCTTTGACCTTATCTTATGATTAAGAAATGCCGAGAG 240
 QY 241 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 300
 Db 241 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 300
 QY 301 GGTGACTTACCTGCTGACAAATGAAGAGAGGCGGCTTTCATGGAAGCTGACTTAA 360
 Db 301 GGTGACTTACCTGCTGACAAATGAAGAGAGGCGGCTTTCATGGAAGCTGACTTAA 360
 QY 361 CCATTCCTTGGCAAAACCGTTAACCAAGACAGATCATCTCTCATTAACCACTCAT 420
 Db 361 CCATTCCTTGGCAAAACCGTTAACCAAGACAGATCATCTCTCATTAACCACTCAT 420
 QY 421 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 480
 Db 421 GGATGGTGTTCGACACTAAGAGAGTGAAGAAATGGAAGTCAATGATGTTGG 480
 QY 481 TATGTATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 Db 481 TATGTATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
 QY 541 GATATATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 Db 541 GATATATATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
 QY 601 CTACC 605
 Db 601 CTACC 605

[illegible]

OY	522	-----AGSGTCTTTATCGCCGGTGATATTATCATGTGAAAGTTTCGCTTAAG	571
Db	712	TCAATGCTTGATTTATTTCTTATCCTACGCCGAGGTAAATCATGCAGAATCATTTCTTTGG	771
OY	572	CTTATAAATATTTAATGAAGGTTTC	598
Db	772	CTAATAAATATTTAATAAATGAAGTTTC	798
RESULT 4			
AXO16322			
LOCUS	AXO16322	576 bp	DNA linear
DEFINITION	Sequence 25 from Patent WO9949046.		PAT 07-SEP-2000
ACCESSION	AXO16322		
VERSION	AXO16322.1	GI:10041878	
KEYWORDS			
SOURCE	Rape:		
ORGANISM	Brassica napus		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
TITLE	Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.		
JOURNAL	1 (bases 1 to 576)		
FEATURES	Robert,J.A., Wyatt,P. and Whitelaw,C. Signal transduction protein involved in plant dehiscence Patent: WO 9949046-A 25 30-SEP-1999; ROBERTS JEREMY ALAN (GB); BIOEMMA UK LTD (GB); WYATT PAUL (GB); WHITELAW CATHERINE (GB) Location/Qualifiers 1..576 /organism="Brassica napus" /db_xref="taxon:3708"		
BASE COUNT	188 a	102 c	122 g 161 t 3 others
ORIGIN			
Query Match	47.8%; Score 289.4; DB: 6; Length 576;		
Best Local Similarity	75.4%; Pred. No. 1,6e-60;		
Matches 416; Conservative 0; Mismatches 94; Indels 42; Gaps 3;			
OY	78	TCGTGATGATGATTCACATPAACCCTTATATATCATGAGAGATCATGAAGCATTTGGG	137
Db	1	TCGCATGATGATTCCTGTAAATACGTAACAATCACAGATATATCATCAAATCAATCGCTG	60
OY	138	GATTTACAGAGAGGAATTAOGGAGGAGGCGGCGATATCATCCACGAGCGGCGGT	197
Db	61	GAATTTACAGAGACTAAAGAACGGTGAAGAGCGATGACATCATCCGCGAGCGCAATG	120
OY	198	CATCTTTGACCTTATCTCATATGATTAAGAAATGCCGAGAGGATGTTTTGCACAA	257
Db	121	CATCTTTGACCTTATCTCATATGATTAAGAAATGCCGAGAGGATGTTTTGCACAA	180
OY	258	CTAAGAGCTAAAGASAAATGSAGAGAGTCATATGTTGGGGTAGCTCACTGCGTC	317
Db	181	CTAAGAGCTTAAGASAAATGAAGTAGACGCTCATGATTAATTTGGGGTAGCAGACTGGCTG	240
OY	318	ACAATGAAGAGAGCGCGAGGGCTTCATATGAGGAAGTCGACTTAACCATTTGCTGGCAAAAC	377
Db	241	ACAATGAAGAGAGCAACTTAAGGCTTCATATGAGGAAGTCGACTTAACCATTTGCTGGCAAAAC	300
OY	378	CGTTAACCAAGAGCAAGATCATCCCTGCATTAACCAACTCATGAGGCTGTAATGG---	434
Db	301	CGTTAACCAAGCGAAGATCTCCCTGCTCATACAACTCATATGATGATGATTTGAATGGATGG	360
OY	435	-----TATATATTTTATATTTATGAAAACACACATATAACGT	471
Db	361	ATGAATTTGCCCATACATATCTCATATTAACAAATGAAAAACACATATATAACGT	420
OY	472	-----CTAAGTGSTATATGATGATGATGATCTGATGATGTTGGATTTAGATTAAG	524
Db	421	CATACACCTGTGTGTATGATGATGATGATCTATCCGCAATGTGTTTTTAAAGGTTGTAT	480
OY	525	GTT-----CTTATGCTCCGATATATAATCATGATGAAGTTGTTGCTTTAAGCTTA	575

DB 481 GTTATTTTATGTGCGTGGCTGATATCAATCANGTAGTCGTTGCTTA 540
OY 576 TAAATATTTAA 587
DB 541 TAAATATTTAA 552

RESULT 5
ATAC016829
LOCUS ATAC016829 110804 bp DNA linear PLN 24-JAN-2001
DEFINITION Arabidopsis thaliana chromosome III BAC T6K12 genomic sequence,
complete sequence.
AC016829
AC016829.7 GI:12408745
KEYWORDS HIG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 110804)
Lin, X., Kaul, S., Town, C.D., Beutler, M., Creasy, T.H., Haas, B.,
Wu, D., Konning, C.M., Koo, H., Fujii, C.Y., Utterback, T.R.,
Barnstead, M.E., Bowman, C.L., White, O., Mierman, W.C. and Fraser, C.M.
Arabidopsis thaliana chromosome III BAC T6K12 genomic sequence
Unpublished
2 (bases 1 to 110804)
Lin, X. and Kaul, S.
Direct Submission
Submitted (08-DEC-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
3 (bases 1 to 110804)
Lin, X.
Direct Submission
Submitted (24-JUN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jan 24, 2001 this sequence version replaced gi:12280893.
Address all correspondence to: at@tigr.org

COMMENT
BAC clone T6K12 is from Arabidopsis chromosome III and is near the
molecular marker ml172.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GAIL (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), GenScan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCAN.html), and Neplantcne
(http://www.cbs.dtu.dk/netgene/cbsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/tcd/at.ac.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html).
Location/Qualifiers
1. 110804
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="III"
/map="ml172"
/clone="T6K12"
complement(join(<1..100,188..556,813..>1275))
/gene="T6K12.1"
complement(<1..>1275)

CDS
/gene="T6K12.1"
complement(join(<1..100,188..556,813..1275))
/gene="T6K12.1"
/note="hypothetical protein, 3' partial"
/codon_start=1
/protein_id="AAF26777.1"
/db_xref="GI:6721150"
/translation="MVHLISLTOTLAIILSLPSIINTSOLDYDTLVFKQCPDLA
NIIQKATIKSPNYSNONLIRQAQLSFLRKLSSESSRSKFLKTLVGEKHAVSGWQ
CREDYSEICHKVGDIREISRSQSGMATSARILRGCHLIVKPERIDTPGAOVNHH
KYKLETPHGLIKHICDQATAEFPPEGEMRETAITAAEGVVDGHGFDYSKLLH
VWQDCGHVADCCGECISAAAIAAEQRMISIAQOILLEGQHVYTHPHLEPNDSY
HESKYNITKSLAIYVGVAALVPAIFMFLKS"
<2503..>3566
/gene="T6K12.2"
<2505..>3566
/gene="T6K12.2"
2505..3566
/gene="T6K12.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAF26778.1"
/db_xref="GI:6721150"
/translation="MGRLEFACPVLEINLISAQDIAPYRNMKYTSVAMINTDPMRL
TTRVDSNANPIMNEKEFVFNDKILYDASAIYEIYAAAKDAIVGVNVLSD
LEAPSGFGDNDGGGNNMMLVTQLIRPRLQGLFRLGVALKIDGQSMPLSIE
VFDGSKRGSRNSYKRDQEA SRMHRTSDQDITLTINDYGVKGVYTGNGEYVYV
AYLLIVRIYIMAYRFAILLDQRLMLRLRKQGTIRKQRLKRIYRYRYRLKXK
QRELSTELNGEDQRKARVERERLDQAAILNGKEERGERGGERRNNNGETVEZ
RKVGGEFVLETLIDVFRP"
3433..3510
/rpt_family="(GGA)n"
4038..4075
/rpt_family="(GA)n"
join(<4235..4346,5141..6750)
/gene="T6K12.3"
<4235..6750
/gene="T6K12.3"
join(4235..4346,5141..6732)
/gene="T6K12.3"
/note="unknown protein"
/codon_start=1
/protein_id="AAF26779.1"
/db_xref="GI:6721151"
/translation="MFGCDFYWSRGISLSESESPRPSLAPLPSPQKGFATG
RILGLIEVYKIKFHRVWSSSHSKSRATFYRADIPGEFHGLGHYCOPTQPLR
GYVLARTSKAVNADPPLKRPVSYSLVWSSSKNGGFWLPPNPGYRAGVIV
THRGPEPEVAVCREDLTESCESEMTLEVSSKSKSGSSPSVSTPCERGM
SOGVAVGPECCYDLSRRYDPDGGCKNDPTLHAMPNLDQVAVTEHRRGPTVYFH
PEYAVHPSYQNFYKNGALILYSKSEQPISTGSNLPAGCCNMDMTIDLPDEBA
KSVLKGNLSSSLYHVPAJAGFTDVTVMVFCFPGPATLKGTLTLPMTIRGEH
VGMEHETRICNFGSELQWFEFSHGSGGWADADIEFVQNKRAVYSKSHGHSFP
HPMVFQGSKLGIGYRNVAKSKYIVVSGGVYVAEYLDKGAVIDPCWLVQVREN
GPVIAVDSGEIKNILNLLPVVRESIEIVDLPFIALYGEEGPGCPREKKNWGEDM
C"
join(7363..8357,8709..9180,9625..9825,9900..10193,
10275..10544,10666..10867,10962..11135,11282..11403,
11540..11741,11949..12077,12175..12663,12753..12827,
13027..13086,13282..13389,13693..13761)
/gene="T6K12.4"
<7363..>13761
/gene="T6K12.4"
/note="C-term similar to cell division protein FTSH
GB:128691 [Escherichia coli]"
join(7363..8357,8709..9180,9625..9825,9900..10193,
10275..10544,10666..10867,10962..11135,11282..11403,
11540..11741,11949..12077,12175..12663,12753..12827,
13027..13086,13282..13389,13693..13761)
/gene="T6K12.4"
/note="unknown protein"
/codon_start=1
/protein_id="AAF26780.1"
/db_xref="GI:6721152"

[illegible]

Db 2947 AATATCAGAGAAAAATTAATAATCTCAAGGTTATTGCTTAACAGCAAAAGCCTT 3006

QY 322 TGAAGAGAGCGAGGCGCTTTCATGAGAGCTGAGCTTAACCATGCTTGCGAAACCGTT 381

Db 3007 TCAAAAAGCAGCATGATTAACCTTTTAAAGAGAGAGCAACTATGCTTAACTAAGCAATT 3066

QY 382 AACCAAGACAGATCATCCCTCTCATTAACCAACATGATGTC 426

Db 3067 ACAAGAAATGATTCTTCTGGCGCATCAACATATCAAGATGC 3111

RESULT 9
AL596454 111752 bp DNA linear PRI 18-DEC-2001
LOCUS Human DNA sequence from clone Rpl1-17718 on chromosome 9, complete
DEFINITION sequence.
ACCESSION AL596454
VERSION AL596454.8 GI:17939740
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Bates, K.
TITLE Direct Submission
JOURNAL Submitted (18-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
On Dec 19, 2001 this sequence version replaced gi:11784505.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: EM: EMBL; S: S;
SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/BSG/Chr9
Rpl1-17718 is from the library RPl1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
Rpl1-17718. It may be shorter because we sequence overlapping
sections only once, except for a short overlap
The true left end of clone Rpl1-515J13 is at 109753 in this
sequence.

FEATURES
Source
1. 111752
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="9"
/clone="Rpl1-17718"
/clone_lib="RPl1-11.1"
BASE COUNT 34576 a 19899 c 20215 g 37062 t
ORIGIN
Query Match 8.5%; Score 51.4; DB 9; Length 111752;
Best Local Similarity 56.1%; Pred. No. 0.021;

Matches 97; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 427 TTGATGAGATATATTATTATATATGAAACACATATTAACGCTTAAGTGTATGTA 486

Db 80799 TTGATGAGATATATTATTATATGAAACACATATTAACGCTTAAGTGTATGTA 80858

QY 487 TGCATGATATCTGCATGATGCTGTTTGAATTTAGAGGTTCTTATCGTGGATATA 546

Db 80859 TCGTGATATCATGTGTGTGTGTGTTTTCATATGTAATGCTTATTTTCCGTATTTT 80918

QY 547 TAAATGATGATGTTGCTTTAAGCTTTAAATATTTAAATAAGGTTTC 599

Db 80919 TAAATGATGATGTTTATTTGACATGAGGTAATAAATATTTGATATACAGCTCC 80971

RESULT 10
AE001420/c 15342 bp DNA linear INV 06-NOV-1998
LOCUS Plasmodium falciparum chromosome 2, section 57 of 73 of the
DEFINITION complete sequence.
ACCESSION AE001420 AE001362
VERSION AE001420.1 GI:3845287
KEYWORDS malaria parasite P. falciparum.
SOURCE Plasmodium falciparum
ORGANISM Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE Gardner, M.J., Tetteh, H., Carucci, D.J., Cummings, L.M., Aravind, L.,
1 (bases 1 to 15342)
AUTHORS Koonin, E.V., Shallow, S., Mason, T., Yu, K., Fujii, C., Pederson, J.,
Shen, K., Jing, J., Aston, C., Lai, Z., Schwartz, D.C., Pertea, M.,
Salzberg, S., Zhou, L., Sutton, G.G., Clayton, R., White, O.,
Smith, H.O., Fraser, C.M., Hoffman, S.L. et al.
TITLE Chromosome 2 sequence of the human malaria parasite Plasmodium
falciparum
JOURNAL Science 282 (3391), 1126-1132 (1998)
MEDLINE 99021743
REMARK Erratum: [[published erratum appears in Science 1998 Dec
4;282(5395):1827]]
REFERENCE 2 (bases 1 to 15342)
AUTHORS Gardner, M.J.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
Medical Center Drive, Rockville, MD 20814, USA

FEATURES
Source
1. 15342
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
/chromosome="2"
2946. 3269
/gene="PFB0830w"
2946. 3269
/gene="PFB0830w"
/note="identified by sequence similarity; putative"
/codon_start=1
/product="ribosomal protein S26"
/protein_id="AAC71955.1"
/db_xref="GI:3845288"
/translation="MPKRRNGRGRKHNRGVNPLRSGNRCVPCXKAKRIRIRNI
VPSNRQDKKMSVSTFQLPRLTIKKQCYCSALHSFVVRBQRARRKAKHY
NPSQL"
complement(4112..5671)
/gene="PFB0835c"
complement(4112..5671)
/gene="PFB0835c"
/note="predicted by Glimmer"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC71956.1"
/db_xref="GI:3845289"
/translation="MKRYINKSKALLLRKMSIVKNSKSNQLEIFPALEKTSILO
KEKLYIEPEVEMICSLNETSCPLFTLVKRSIDIVVLEKRVPCGSGKTNMCM
LNNDKTSFTDKGFIINKKEGEERKKNETNMNENKRVDFDSFVQNLIVLDFPK
VFYKHIILIDSLAADVYRHSIGVYRNRKYNRVVLFNNLTKYERKNVHHHFSFA


```

LHDSVPLINCYIVNSGTSYLLKLDPEQONNVFDIYTFVAVNCLFFKPOQVYVPL
NSMONEKQKJIKSLCSHKIDVQSLFENKREKSTLPHNDHEILSNFRKXQONCLYI
KINNEYIYDNEQONELVNSNDNILLIPSLINDNDFSHAKVOYACKRDYVNSVL
FISSEPKINHNFEKGHLAYLSPCHAGVAYKTLPLPIIMQINSLOYIILPVPVNSVL
YKGVYVSDHSIOYVHAYTCFDEKRAYVLTTCDSNFPNSDISFVQSOEERKLY
MLGIRSOILP
7534..8526
/gene="PF08040w"
7534..8526
/gene="PF08040w"
7534..8526
/gene="PF08040w"
/note="identified by sequence similarity; putative"
/codon_start=1
/product="replication factor C, 40 kDa subunit
(replication activator)"
/protein_id="AAC71957.1"
/db_xref="GI:3845290"
/translation="MENIPWYEKYRPRKLDIYHONNAVMLKEVYRKMPHLIFPG
PPGKTSATNALAELEKGENISERYLENASDGGINVRKIKAYRISISKKI
HSETRKVPKRIYVDEADMTEDNQSALRIEYISNTRELLICNYIHKISDPI
SRCSYRPOSIPINIKKELIYICQNDIDYDALKETIETEGDRAVSLDQCS
CINRKLINSVLDSGLPSDNIYKIIDACKMDKLVEYVDDIIEGPDVAYIKRS
FNNFVNTYEDSLKYLILLESRDYRLHGCATYIQLSPASSVSHLSNV"
9688..10611
/gene="PF08045w"
9688..10611
/gene="PF08045w"
/note="predicted by GlimmerM"
/codon_start=1
/product="predicted integral membrane protein"
/protein_id="AAC71958.1"
/db_xref="GI:3845291"
/translation="MDSYIEKRSNVLKNQYDIKIQNEVDELTSYIDKYSSEDM
NNYKTSIDEDYMYOKINDNDNSERDPCINAYIEMTEKREKREKMTKL
ESINCYDFENDFLIKVPEKISNSMKNHNIYKLEFLVLYTSFTFTFR
KYIYNLPEKHNFLSEKVENIKMHRKNSVLPYDRIQYIVNKFEDVPO
FLCLLYTLNGKDFIIFMLYSIIYFVINAHYGLNPLMGLNLFSTYLLLEYI
NKKRKLNIYIYIYICMYFLCIYINFLI"
complement(11963..14386)
/gene="PF08050c"
complement(11963..14386)
/gene="PF08050c"
complement(11963..14386)
/note="predicted by GlimmerM"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC71959.1"
/db_xref="GI:3845292"
/translation="MAVESKPNSSKREKNEENDINKQDSNKNKINGENIPEYKGI
NESGMSNDINKNOEKKKKKKKKNTKKRYNININININININININININININ
ERDNIYINKNITNIDSSYVEGENRNDINNNNNNNININININININININININ
KKITLKRNDIKDEGINNENITTLNKNNDKNNNNNNDNRNNNNNNNNNNNNNN
CCSEKTLQREKEYNIRARLESNFKKQKNOVQTEQNNINHTYLNNNIINNINNGDN
OYAYINNHYHAYHNSYNHAYIRQNNIPICININHPAIEKLNPPYHDIYATVYM
YSTOKNNMKTKQIGHYGINNEDNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
DNYKTKNPNQGTNNFNOGTNNFCTNNENNAKNIKININININININININININ
NVEYENKNIYDANGNINNNISYTLCKNNNNINENIHESPINCOHNTEKYNNDTN
FNEPTNKKKREKRIHNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
VKNTEONLQKRNKQKQSKOSNNKNNNSHKKQINIVTNNMDKNNNSH.SKVI
VDNKLKSHADNSNEIVTKGRKNNKNNKNNKNNINIVANNINNSNNNNNNNN
SMNNVNNNNNNPYPNVNIOQDSNIALLVNKKNNIDPNNFOLNHNHNNNNNITN
NVMNNNNLTTSNENYNNLINSYEPFEYENLMDLYCRDISLYEKYRDGNDLQQHK
RYDIFPPL"
BASE COUNT 6057 a 1268 c 1441 g 6576 t
ORIGIN

```

```

Db 7544 AATATTCACCTTAAATAATTA-AAATATAAATAATATAATATAATATAA 7486
Oy 453 GAAACACATATATACCTCTAAGTGTATGTATGATAGATACTGATGTGTGT 512
Db 7485 TATATATATATATATATATATATATATGATATGATATATATATATATAT 7426
Oy 513 TTAGATTTAGGTTCTTATCGTCGCGATATATATATATATATATATATATAGC 572
Db 7425 TATATATATATATATATATATATATATATATATATATATATATATATATAT 7366
Oy 573 TTATATA 579
Db 7365 AATATA 7359

RESULT 11
PFMA113P2_3/c
Sequence split into 4 fragments LOCUS PFMA113P2 Accession AL049185
Fragment Name Begin End
PFMA113P2_0 1 110000
PFMA113P2_1 100001 210000
PFMA113P2_2 200001 310000
PFMA113P2_3 300001 377835
Continuation (4 of 4) of PFMA113P2 from base 300001 (AL049185 Plasmodium falciparum)

Query Match 8.1%; Score 49.2; DB 2; Length 77835;
Best Local Similarity 51.4%; Pred. No. 0.074;
Matches 114; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

Oy 380 TTAACCAAGGACAAATATACCTCTCATTAACCAACATGATGCTGATGATATAT 439
Db 56551 TGAAGCAGGATACAGCAGACAGACCTTTAAACACCTTGCTCCAGAAATTA 56492
Oy 440 ATTTATATATATAGAACACACATATATATATATATATATATATATATATAT 499
Db 56491 AATATAATATATATATATATATATATATATATATATATATATATATATAT 56432
Oy 500 GCATGCTGTCTTTAGATTTAGGTTCTTATCGCCGATATATATATATATAT 559
Db 56431 GATATATATATATATATATATATATATATATATATATATATATATATAT 56372
Oy 560 TGTGCTTAACCTTAAATAATTTAAATAAGGTTTCCTC 601
Db 56371 TTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTTC 56330

RESULT 12
AC018519 standard; DNA; HTG; 193488 BP.
AC018519;
AC AC018519.4
SV AC018519.4
Xx 15-DEC-1999 (Rel. 62, Created)
Dt 28-MAY-2000 (Rel. 63, Last updated, Version 3)
Xx Homo sapiens chromosome 4 clone RP11-402c9 map 4, WORKING DRAFT SEQUENCE,
DE 25 unordered pieces.
Xx
Xx HTG; HTGS_DRAFT; HTGS_PHASE1.
Xx
Xx Homo sapiens (human)
Os Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
Oc Eutheria; Primates; Catarrhini; Homindae; Homo.
Xx
Xx [1]
Xx 1-193488
Ra Birren B., Linton L., Nusbaum C., Lander E.;
Rt "Homo sapiens chromosome 4, clone RP11-402c9";
Rt Unpublished.

```


XX [2]
 RN 1-193488
 RP Birren B., Linton L., Nusbaum C., Lander E., Abraham H., Allen N.,
 RA Anderson S., Baldwin J., Barna N., Beckerly R., Beda F., Boguslavsky L.,
 RA Boukagater B., Brown A., Castle A., Colangelo M., Collins S.,
 RA Collymore A., Cooke P., Dekrellan K., Dewar K., Domino M., Doyle M.,
 RA Fennel J., Ferreira P., Fitzhugh W., Forrest C., Gage D., Galagan J.,
 RA Gardna S., Grant G., Hago B., Heatford A., Horton L., Howland J.,
 RA Johnson R., Jones C., Kann L., Karatas A., Klein J., Landers T.,
 RA Lehoczy J., Liu C., Liu G., Locke K., Macdonald P., Margulis N.,
 RA McEwen P., McGurk A., McKenna K., Meldrum J., Morrow J., Naylor J.,
 RA Norman C.H., O'Connor T., O'Donnell P., Peterson K., Pierre N., Pollara V.,
 RA Riley R., Rottman D., Roy A., Santos R., Severy P., Stange-Thomann N.,
 RA Stojanovic N., Subramanian A., Talamas J., Vestly S., Theodore J.,
 RA Tirrell A., Vassiliev H., Viel R., Vo A., Wu X., Wyman D., Ye W.J.,
 RA Zimmer A., Zody M.;
 RT Submitted (13-DEC-1999) to the EMBL/Genbank/DBJ databases.
 RL Whitehead Institute/MIT Center for Genome Research, 320 Charles Street,
 RL Cambridge, MA 02141, USA
 RL
 XX On May 26, 2000 this sequence version replaced gi:6649474.
 CC All repeats were identified using RepeatMasker:
 CC Smit, A.F.A. & Green, P. (1996-1997)
 CC http://ftp.genome.washington.edu/RM/RepeatMasker.html
 CC ----- Genome Center
 CC Center: Whitehead Institute/ MIT Center for Genome Research
 CC Center code: WIBR
 CC Web site: http://www.seq.wi.mit.edu
 CC Contact: sequence_submissions@genome.wi.mit.edu
 CC ----- Project Information
 CC Center project name: L2439
 CC -----
 CC Center clone name: 402.C.9
 CC ----- Summary Statistics
 CC Sequencing vector: MJ3; M77815; 100% of reads
 CC Chemistry: Dye-terminator Big Dye; 100% of reads
 CC Assembly program: Phrap; version 0.960731
 CC Consensus quality: 177768 bases at least Q30
 CC Consensus quality: 185018 bases at least Q40
 CC Consensus quality: 187917 bases at least Q20
 CC Insert size: 182000; agarose-fp
 CC Insert size: 191088; sum-of-ctfils
 CC Quality coverage: 4.7 in Q20 bases; sum-of-ctfils
 CC Quality coverage: 4.5 in Q20 bases; sum-of-ctfils
 CC -----
 CC * NOTE: This is a 'working draft' sequence. It currently
 CC * consists of 25 contigs. The true order of the pieces
 CC * is not known and their order in this sequence record is
 CC * arbitrary. Gaps between the contigs are represented as
 CC * runs of N, but the exact sizes of the gaps are unknown.
 CC * This record will be updated with the finished sequence
 CC * as soon as it is available and the accession number will
 CC * be preserved.
 CC *
 CC 1 1339: contig of 1339 bp in length
 CC * 1340 1439: gap of 100 bp
 CC * 1440 2663: contig of 1224 bp in length
 CC * 2664 2763: gap of 100 bp
 CC * 2764 4211: contig of 1448 bp in length
 CC * 4212 4311: gap of 100 bp
 CC * 4312 5618: contig of 1307 bp in length
 CC * 5619 5718: gap of 100 bp
 CC * 5719 6795: contig of 1077 bp in length
 CC * 6796 6895: gap of 100 bp
 CC * 6896 9287: contig of 2392 bp in length
 CC * 9288 9387: gap of 100 bp
 CC * 9388 11178: contig of 1791 bp in length
 CC * 11179 11278: gap of 100 bp
 CC * 11279 13748: contig of 2470 bp in length
 CC * 13749 13848: gap of 100 bp
 CC * 13849 15939: contig of 2091 bp in length
 CC * 15940 16039: gap of 100 bp
 CC * 16040 17849: contig of 1810 bp in length

CC * 17850 17949: gap of 100 bp
 CC * 17950 20794: contig of 2845 bp in length
 CC * 20795 20894: gap of 100 bp
 CC * 20895 23299: contig of 2405 bp in length
 CC * 23300 23399: gap of 100 bp
 CC * 23400 30423: contig of 7024 bp in length
 CC * 30424 30523: gap of 100 bp
 CC * 30524 36377: contig of 5854 bp in length
 CC * 36378 36477: gap of 100 bp
 CC * 36478 45186: contig of 8709 bp in length
 CC * 45187 45286: gap of 100 bp
 CC * 45287 53168: contig of 7882 bp in length
 CC * 53169 53268: gap of 100 bp
 CC * 53269 62404: contig of 9136 bp in length
 CC * 62405 62504: gap of 100 bp
 CC * 62505 73127: contig of 10623 bp in length
 CC * 73128 73227: gap of 100 bp
 CC * 73228 82813: contig of 9586 bp in length
 CC * 82814 82913: gap of 100 bp
 CC * 82914 94881: contig of 11968 bp in length
 CC * 94882 94981: gap of 100 bp
 CC * 94982 107947: contig of 12966 bp in length
 CC * 107948 108047: gap of 100 bp
 CC * 108048 121648: contig of 13601 bp in length
 CC * 121649 121748: gap of 100 bp
 CC * 121749 136375: contig of 14627 bp in length
 CC * 136376 136475: gap of 100 bp
 CC * 136476 1359593: contig of 23118 bp in length
 CC * 1359594 159693: gap of 100 bp
 CC * 159694 193488: contig of 33795 bp in length.
 CC
 CC XX
 FH Key
 FH Location/Qualifiers
 FT source
 FT 1. 193488
 FT /chromosome="4"
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /map="4"
 FT /clone="RP11-402C9"
 FT /clone_1lb="RPC1-11 Human Male BAC"
 FT 1. 1339
 FT misc-feature
 FT /note="assembly-fragment"
 FT 1440. 2663
 FT /note="assembly-fragment"
 FT 2764. 4211
 FT /note="assembly-fragment"
 FT 4312. 5618
 FT /note="assembly-fragment"
 FT 5719. 6795
 FT /note="assembly-fragment"
 FT 6896. 9287
 FT /note="assembly-fragment"
 FT 9388. 11178
 FT /note="assembly-fragment"
 FT 11279. 13748
 FT /note="assembly-fragment"
 FT 13849. 15939
 FT /note="assembly-fragment"
 FT 16040. 17849
 FT /note="assembly-fragment clone_end:17 vector_side:left"
 FT 17950. 20794
 FT /note="assembly-fragment"
 FT 20895. 23299
 FT /note="assembly-fragment"
 FT 23400. 30423
 FT /note="assembly-fragment"
 FT 30524. 36377
 FT /note="assembly-fragment"
 FT 36478. 45186
 FT /note="assembly-fragment"
 FT 45287. 53168
 FT /note="assembly-fragment"
 FT 53269. 62404
 FT /note="assembly-fragment"
 FT 62505. 73127
 FT /note="assembly-fragment"
 FT 73228. 82813
 FT /note="assembly-fragment"
 FT 82914. 94881
 FT /note="assembly-fragment"
 FT 94982. 107947
 FT /note="assembly-fragment"
 FT 108048. 121648
 FT /note="assembly-fragment"
 FT 121749. 136375
 FT /note="assembly-fragment"
 FT 136476. 1359593
 FT /note="assembly-fragment"
 FT 1359594. 159693
 FT /note="assembly-fragment"
 FT 159694. 193488
 FT /note="assembly-fragment"

```

FT      misc_feature      /note="assembly-fragment"
FT      62505..73127
FT      /note="assembly-fragment"
FT      7328..82813
FT      misc_feature      /note="assembly-fragment"
FT      82914..94881
FT      /note="assembly-fragment"
FT      94982..107947
FT      misc_feature      /note="assembly-fragment clone_end:SP6 vector_side:left"
FT      108048..121648
FT      /note="assembly-fragment"
FT      121749..136375
FT      misc_feature      /note="assembly-fragment"
FT      136476..159593
FT      /note="assembly-fragment"
FT      159694..183488
FT      misc_feature      /note="assembly-fragment"
XX      SQ      Sequence 193488 BP; 57106 A; 37991 C; 37536 G; 58449 T; 2406 other;

Query Match      8.1%; Score 49; DB 30; Length 193488;
Best Local Similarity 56.5%; Pred. No. 0.08; Mismatches 70; Indels 0; Gaps 0;
Matches 91; Conservative 0;

QY      430 ATGGATATATTTATTTATTTATGGAACACACATATTAAGCTCTAGTGTATATGC 489
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
DB      30234 ATATATATATATATATATATATACACACACACACACAAATATATATATATATAT 30293
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
QY      490 ATAGATCTTCATGCTGTGTGTTTGAATTTAGGTTCTTATCGCCGATATATTA 549
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
DB      30294 ACACATACATGCTGTGTATATATACACATATATGCTGTATATATATACACATATAC 30353
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
QY      550 TCATGTAGTGTGCTTTAAGCTTATTAATAATAATAATAATAATAATAATAATAATA 590
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
DB      30354 ACATATATATTTGTATATATACACATATATATGTAATA 30394
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||

RESULT 13
AC108142/c 198509 bp DNA linear HTG 08-FEB-2002
LOCUS      Homo sapiens chromosome 4 clone RP11-402C9, WORKING DRAFT SEQUENCE,
DEFINITION 8 unordered pieces.
AC108142 AC018519
VERSION    AC108142.2 GI:18640709
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVERIN.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 198509)
AUTHORS   Waterston,R.H.
TITLE     The sequence of Homo sapiens clone
JOURNAL   Unpublished
AUTHORS   2 (bases 1 to 198509)
REFERENCE Waterston,R.H.
AUTHORS   Direct Submission
TITLE     Submitted (25-JAN-2002) Genome Sequencing Center, Washington
JOURNAL   University School of Medicine, 4444 Forest Park Parkway, St. Louis,
AUTHORS   MO 63108, USA
COMMENT    On Feb 8, 2002 this sequence version replaced gi:18370048.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
Project Information
Center project name: H_NH0402C09
Drafting center: MIBR
----- Summary Statistics -----
Sequencing vector: M13; 35%
Sequencing vector: plasmid; 65%

```

```

Chemistry: Dye-Primer ET; 0% of reads
Chemistry: Dye-Terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 192369 bases at least Q40
Consensus quality: 194554 bases at least Q30
Consensus quality: 195481 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 197809; sum-of-contigs
Quality coverage: 13.76 in Q20 bases; sum-of-contigs
Quality coverage: 13.76 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1407: contig of 1407 bp in length
* 1408: gap of unknown length
* 1508: contig of 1329 bp in length
* 2836: gap of unknown length
* 2937: gap of unknown length
* 4636: contig of 1700 bp in length
* 4637: gap of unknown length
* 4737: contig of 1543 bp in length
* 6280: gap of unknown length
* 6380: contig of 1191 bp in length
* 7571: gap of unknown length
* 7671: contig of 2907 bp in length
* 10578: gap of unknown length
* 10678: contig of 2502 bp in length
* 13179: gap of unknown length
* 13279: gap of unknown length
* 13280: 198509: contig of 185230 bp in length.
*
FEATURES
Source
1..198509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/clone="RP11-402C9"
1..1407
/note="assembly-name:Contig10"
1508..2836
/note="assembly-name:Contig24"
2937..4636
/note="assembly-name:Contig27"
4737..6279
/note="assembly-name:Contig29"
6380..7570
/note="assembly-name:Contig34"
7671..10577
/note="assembly-name:Contig33"
10678..13179
/note="assembly-name:Contig34"
13280..198509
/note="assembly-name:Contig35"
BASE COUNT 59384 a 38737 c 39643 g 60041 t 704 others
ORIGIN
Query Match      8.1%; Score 49; DB 2; Length 198509;
Best Local Similarity 56.5%; Pred. No. 0.07;
Matches 91; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY      430 ATGATATATTTATTTATTTATGGAACACACATATTAAGCTCTAGTGTATATGC 489
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
DB      127658 ATATATATATATATATATATATACACACACACAAATATATATATATATATAT 127599
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
QY      490 ATAGATCTTCATGCTGTGTGTTTGAATTTAGGTTCTTATCGCCGATATATTA 549
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
DB      127598 ACACATACATGCTGTGTATATATACACATATATGCTGTATATATATACATATATAC 127539
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||
QY      550 TCATGTAGTGTGCTTTAAGCTTATTAATAATAATAATAATAATAATAATAATAATA 590
      || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| ||

```

Db 127538 ACATATATTTGTGTATATACACATATATATGTGTATA 127498

RESULT 14
AX016326 1324 bp DNA linear PAT 07-SEP-2000
LOCUS
DEFINITION Sequence 29 from Patent WO9949046.
ACCESSION AX016326
VERSION AX016326.1 GI:10041881
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1324)
REFERENCE
AUTHORS Roberts, J.A., Wyatt, P. and Whitelaw, C.
TITLE Signal transduction protein involved in plant dehiscence
JOURNAL Patent: WO 9949046-A 29 30-SEP-1999;
ROBERTS JEREMY ALAN (GB); BIOGEMMA UK LTD (GB); WYATT PAUL (GB);
WHITELAW CATHERINE (GB)
FEATURES
SOURCE location/Qualifiers
1..1324
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
1200..>1322
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC07687.1"
/db_xref="GI:10041882"
/translation="MATKSTGCTGCTKSEVKKLIINLVYDDPLNRLHECHQ"
BASE COUNT 460 a 214 c 182 g 467 t 1 others
ORIGIN

Query Match 8.1%; Score 48.8; DB 6; Length 1324;
Best Local Similarity 71.3%; Pred. No. 0.11; Mismatches 17; Indels 18; Gaps 1;
Matches 87; Conservative 0;

QY 12 AATGAGATGGCAACAAATCCATGGAGATTCGAGAAATTAAGAAACTA----- 67
Db 1192 AATGAGATGGCAACAAATCCATGGAGATTCGAGAAATTAAGAAACTA----- 67
QY 68 -----AACGTGTGATCGTCGATGATGTCACATTAACCTTATATTCATG 113
Db 1252 AGAGAAACTAATCAAGCTGTATGCTGCATATATATCATTAACCTTACACTCCACG 1311
QY 114 AG 115
Db 1312 AG 1313

RESULT 15
AC099331 151498 bp DNA linear PRI 12-FEB-2002
LOCUS
DEFINITION Homo sapiens chromosome 3 Clone RP11-90B15, complete sequence.
ACCESSION AC099331 AC073441
VERSION AC099331.2 GI:18652541
KEYWORDS
SOURCE HTG.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 151498)
REFERENCE
AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Unpublished
2 (bases 1 to 151498)
AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
REFERENCE 3 (bases 1 to 151498)

AUTHORS Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
Saenphimachak, C., Phelps, K.A., Raymond, C. and Haugen, E.D.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-2002) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
COMMENT On Feb 12, 2002 this sequence version replaced gi:16674865.
----- Genome Center
Center: University of Washington Genome Center
Center Code: UWGC
Web site: <http://www.genome.washington.edu>
Contact: uwgchgs@u.washington.edu
Drafting Center: BCM
----- Project Information
Center project name: chr-3
Center clone name: RP11-90B15 (bc0621)
----- Summary Statistics
Sequencing vector: unknown; 3% of reads
Sequencing vector: Plasmid; 68% of reads
Chemistry: Dye-terminator ET; 9% of reads
Chemistry: Dye-terminator Big Dye; 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151498 bases at least Q40
Consensus quality: 151494 bases at least Q30
Consensus quality: 151498 bases at least Q20
Insert size: 149290; sum-of-contigs
Quality coverage: 9.0x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': Mapping in progress
3': RP11-613N24 (UWGC:bc0491) AC099557

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Bases by base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

HindIII	BglII	EcoRI
SeqDerMap FngPrint	SeqDerMap FngPrint	SeqDerMap FngPrint
-----	-----	-----
970	964	8580
-----	-----	-----
6382	6463	2067
-----	-----	-----
512	<800	5474
-----	-----	-----
449	<800	6597
-----	-----	-----
	6842	5809
-----	-----	-----
		5961

Claim 5; Fig 1; 71pp; English.

-
PT
XX
PS
XX
Claim 5; Fig 1; 71pp; English.
CC The invention provides a nucleic acid encoding a signal transduction
CC protein involved in the process of dehiscence. The nucleic acids and
CC proteins are useful for regulating or controlling dehiscence of a pod or
CC an anther in a plant, useful in the production of male sterile plants.
CC The methods, etc., may be used in production of shatter resistance or
CC shatter-delayed plants such as oilseed rape (*Brassica napus*). The present
CC sequence represents the DNA encoding a B. napus response regulator
CC protein DZ2.

Sequence 605 BP; 198 A; 101 C; 140 G; 166 T; 0 other;

Query Match	100.0%	Score 605;	DB 20;	Length 605;
Best Local Similarity	100.0%	Pred. No. 3.9e-159;		
Matches 605; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

[illegible]

XX Signal transduction protein; dehiscence; male sterile plant; DZ2 gene;
KW shatter resistance; oilseed rape; response regulator protein; ss.

OS Brassica napus.

PN WO9949046-A1.

PD 30-SEP-1999
YY

PE 22-MAR-1999; 99WC-GB00905.
XX

PR 20-MAR-1996; 90GB-00006113.
XX
XX

XX
XX
XX

(P) 00 / 2400000000 ON ELD

XX XX

DR P-PSDB; AAY42653.

PT A nucleic acid er

PS Example 2; Fig 6; 71pp; English.

CC The invention provides a nucleic acid encoding a signal transduction
CC protein involved in the process of dehiscence. The nucleic acids and
CC proteins are useful for regulating or controlling dehiscence of a pod or
CC an anther in a plant, useful in the production of male sterile plants.
CC The methods, etc., may be used in production of shatter resistance or
CC shatter-delayed plants such as oilseed rape (*Brassica napus*). The present
CC sequence represents the B. napus D22 sequence.

Sequence 605 BP; 198 A; 101 C; 140 G; 166 T; 0 other;

Query Match	100.0%	Score 605;	DB 20;	Length 605;
Best Local Similarity	100.0%	Pred. No. 3.9e-15;		
Matches 605; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1	GGCCACGACGAGATGGAAGATGGCAACAAAATCCATGGGAGATATCGAGAAAATTAAGAA	60
Db	1	GGGACGAGGAGATGGAAGATGGCAACAAAATCCATGGGAGATATCGAGAAAATTAAGAA	60
QY	61	GAAACTAAACGGTGTGATCGTCGATGATGATCACTAAACCTTATTAATTCATGAGGAAT	120
Db	61	GAAACTAAACGGTGTGATCGTCGATGATGATCACTAAACCTTATTAATTCATGAGGAAT	120
QY	121	CATCAAAAGCGATTGGGGTATTTTCACAGACACGGAATTAAGGTGAGAGGCGATTAATCAT	180
Db	121	CATCAAAAGCGATTGGGGTATTTTCACAGACACGGAATTAAGGTGAGAGGCGATTAATCAT	180
QY	181	CCACGGTGAAGGCGGCTCATCTTTTGAACCTATCTCAATGATTAAGTAAGAAATGGCCGAG	240
Db	181	CCACGGTGAAGGCGGCTCATCTTTTGAACCTATCTCAATGATTAAGTAAGAAATGGCCGAG	240
QY	241	GGATGGTGTGTTGACAACTAAGAAAGCTAAGAGAAATGAGAGTGAAGTCATGATTGTGG	300
Db	241	GGATGGTGTGTTGACAACTAAGAAAGCTAAGAGAAATGAGAGTGAAGTCATGATTGTGG	300
QY	301	GGGATTTACAGTGGCTGACAAATGAAGAGAGCCGAGGCTTCATGAGAAAGTGGACCTTA	360
Db	301	GGGATTTACAGTGGCTGACAAATGAAGAGAGCCGAGGCTTCATGAGAAAGTGGACCTTA	360
QY	361	CCATTGCTGGGCAAAACGGTTAACCAAGGACAGATCATCCCTCATTAACCAACTCAT	420
Db	361	CCATTGCTGGGCAAAACGGTTAACCAAGGACAGATCATCCCTCATTAACCAACTCAT	420
QY	421	GGATGCTTATGATATATATTTATATTTATGAAAACACATTAATTAAGCTTAAGTGTG	480
Db	421	GGATGCTTATGATATATATTTATATTTATGAAAACACATTAATTAAGCTTAAGTGTG	480

QY 481 TATGTATGATAGATAGTTCATGTCGTGTTTAAATTTAGGGTCTTATCGCCGT 540
D 481 TATGTATGATAGATAGTTCATGTCGTGTTTAAATTTAGGGTCTTATCGCCGT 540
QY 541 GATATATATCATGTTAGTGTGCTTTAGCTTATATAAATTTAAATAGGTTTCT 600
D 541 GATATATATCATGTTAGTGTGCTTTAGCTTATATAAATTTAAATAGGTTTCT 600
QY 601 CTACC 605
D 601 CTACC 605
RESULT 3
AAC35077
ID AAC35077 standard; DNA: 834 BP.
AAC35077;
AC
XX
XX
DT 17-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 8915.
XX
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135553.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147702.
PR 05-AUG-1999; 99US-0147792.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141847.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142927.
PR 14-JUL-1999; 99US-0143544.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144871.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 23-JUL-1999; 99US-0145122.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145228.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145518.
PR 28-JUL-1999; 99US-0145519.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158222.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160757.
PR 21-OCT-1999; 99US-0160758.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162122.

Query Match 49.1%; Score 296.8; DB 21; Length 700;
Best Local Similarity 81.5%; Pred. No. 4.6e-73;
Matches 362; Conservative 0; Mismatches 62; Indels 18; Gaps 1;

QY 12 AATGAGATGCGCAAAATCCATGGAGATATCCAGGAATAAAGAAACTA----- 67
DB 88 AATCGAATGCAACAAATCCCGAGAGTACCGAGAAACCAAGTGAAGAGTGA 147
QY 68 -----AACGTGTTGATCGTCGATGATGATCCACTAACTTAATATTCATG 113
DB 148 AGAAGAACTATATCAAGCTGTGATGTCGATGATGATCATTAACCTAGACTCAG 207
QY 114 AGAAGATCATCAAGCATTTGGGGTTTTCACAGACAGCAATAACGGTGAAGACG 173
DB 208 AGATGATCATCAAAACATCGAGGATTTCTGACTCAAGAAAGATGCGAGAGCGAG 267
QY 174 TAATCATCAACCGTGAAGCGGCTCACTTTGACCTTAATGATGATAAGAAATGC 233
DB 268 TGATCTCCACACCGTGAAGCGGAGCATCTTCCACCTTATCTTAAGATAAGAAATGC 327
QY 234 CCGAGAGGATGTGTTTGCACAACATAAGAACTAAGAAATGAAGTGAAGTGAATGA 293

PR	03-JUN-1999	9905-01372228
PR	03-JUN-1999	9905-01375758
PR	04-JUN-1999	9905-01375022
PR	07-JUN-1999	9905-01377724
PR	08-JUN-1999	9905-01380904
PR	10-JUN-1999	9905-01385400
PR	10-JUN-1999	9905-01388847
PR	18-JUN-1999	9905-01394577
PR	18-JUN-1999	9905-01394548
PR	14-JUN-1999	9905-01391915
PR	16-JUN-1999	9905-01394522
PR	16-JUN-1999	9905-01394543
PR	17-JUN-1999	9905-01394542
PR	18-JUN-1999	9905-01394541
PR	18-JUN-1999	9905-01394622
PR	18-JUN-1999	9905-01394633
PR	18-JUN-1999	9905-01394655
PR	18-JUN-1999	9905-01397500
PR	18-JUN-1999	9905-01397633
PR	21-JUN-1999	9905-01398177
PR	22-JUN-1999	9905-01398789
PR	22-JUN-1999	9905-01403533
PR	23-JUN-1999	9905-01403534
PR	22-JUN-1999	9905-01403543
PR	24-JUN-1999	9905-01406835
PR	24-JUN-1999	9905-01406835
PR	28-JUN-1999	9905-01408233
PR	28-JUN-1999	9905-01409911
PR	30-JUN-1999	9905-01412677
PR	01-JUL-1999	9905-01418442
PR	01-JUL-1999	9905-01436244
PR	01-JUL-1999	9905-01421055
PR	01-JUL-1999	9905-01420545
PR	06-JUL-1999	9905-01423800
PR	08-JUL-1999	9905-01428033
PR	08-JUL-1999	9905-01429200
PR	12-JUL-1999	9905-01432977
PR	12-JUL-1999	9905-01435424
PR	15-JUL-1999	9905-01436244
PR	15-JUL-1999	9905-01440055
PR	15-JUL-1999	9905-01440855
PR	15-JUL-1999	9905-01440866
PR	15-JUL-1999	9905-01443235
PR	15-JUL-1999	9905-01443321
PR	15-JUL-1999	9905-01443331
PR	15-JUL-1999	9905-01443333
PR	19-JUL-1999	9905-01443334
PR	19-JUL-1999	9905-01443335
PR	19-JUL-1999	9905-01443522
PR	20-JUL-1999	9905-01445322
PR	20-JUL-1999	9905-01446322
PR	20-JUL-1999	9905-01448644
PR	21-JUL-1999	9905-01448184
PR	21-JUL-1999	9905-01450086
PR	21-JUL-1999	9905-01450086
PR	22-JUL-1999	9905-01450085
PR	22-JUL-1999	9905-01450877
PR	22-JUL-1999	9905-01450899
PR	22-JUL-1999	9905-01451922
PR	22-JUL-1999	9905-01451545
PR	23-JUL-1999	9905-01452118
PR	23-JUL-1999	9905-01452244
PR	26-JUL-1999	9905-01452761
PR	27-JUL-1999	9905-01452913
PR	27-JUL-1999	9905-01455918
PR	27-JUL-1999	9905-01455919
PR	28-JUL-1999	9905-01455911
PR	02-AUG-1999	9905-01463866
PR	02-AUG-1999	9905-01463868
PR	02-AUG-1999	9905-01463899
PR	03-AUG-1999	9905-01467038
PR	04-AUG-1999	9905-01472044
PR	04-AUG-1999	9905-01473022


```

XX      Wyatt P, Roberts JA, Whitelaw C;
PI      WPI: 1999-580449/49.
XX      P-PSDB: AAY42652.
DR      A nucleic acid encoding a signal transduction protein involved in plant
XX      dehiscence, useful for producing shatter resistant male sterile plants
PT      -
PS      Example 2; Fig 6; 71pp; English.
XX      The invention provides a nucleic acid encoding a signal transduction
XX      protein involved in the process of dehiscence. The nucleic acids and
XX      proteins are useful for regulating or controlling dehiscence of a pod or
XX      an anther in a plant, useful in the production of male sterile plants.
XX      The methods, etc. may be used in production of shatter resistance or
XX      shatter delayed plants such as oilseed rape (Brassica napus). The present
XX      sequence represents the B. napus DZ2B cDNA sequence.
SQ      Sequence 576 BP; 188 A; 102 C; 122 G; 161 T; 3 other;

Query Match          47.8%; Score 289.4; DB 20; Length 576;
Best Local Similarity 75.4%; Pred. No. 5e-71;
Matches 416; Conservative 0; Mismatches 94; Indels 42; Gaps 3;

QY      78 TCCTCATATGATCCACTTAACTTATATCATGAGAGATCATCAAGCATTTGGGG 137
DB      1 TCGTCNATGATGATCCTGTAATACGTAACCTACAGAGATATCATCAATCAATCGGTG 60
QY      138 GTATTTACAGACGAGCAATTAAGGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 197
DB      61 GAATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY      198 CATCTTTGACCTTATCTTAATGATTAAGAAATGCCGAGAGAGAGAGAGAGAGAGAGAG 257
DB      121 CATCTTGCACCTTATCTTAATGATTAAGAAATGCCGAGAGAGAGAGAGAGAGAGAGAG 180
QY      258 CTAAAGAGCTTAAGAGAAATGGAAGTGAATGATGTTGGGGTCACTTCAGTGGGTG 317
DB      181 CTAAAGAGCTTAAGAGAAATGGAAGTGAATGATGTTGGGGTCACTTCAGTGGGTG 240
QY      318 ACAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 377
DB      241 ACAATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
QY      378 CGTTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 434
DB      301 CGTTAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
QY      435 -----TATATTTTATATATGAAACACACATTAATACGT 471
DB      361 ATGAATGGCCACTACATATCTACATTAACAATATGAAAAACACATTAATTAACGT 420
QY      472 -----CTAAGTGTATATGATGATGATGATGATGATGATGATGATGATGATG 524
DB      421 CATACACCTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 480
QY      525 GTT-----CTTATGTCGCGATATATATATATATATATATATATATATATATAT 575
DB      481 GTTGTATTTTATTTGTCGCGATATATATATATATATATATATATATATATATATAT 540
QY      576 TAAATATATTTAA 587
DB      541 TAAATATATGAA 552

```

RESULT 8

```

AA222975
ID      AA222975 standard; DNA; 1716 BP.
XX      AC      AA222975;
XX

```

```

DT      10-JAN-2000 (first entry)
XX      Brassica napus DZ2B promoter region and partial coding sequence.
DE      Signal transduction protein; dehiscence; male sterile plant; DZ2B gene;
XX      Shatter resistance; oilseed rape; response regulator protein; ss.
XX      Brassica napus.
OS      WO949046-A1.
XX      30-SEP-1995.
XX      22-MAR-1999; 99WO-GB00905.
XX      20-MAR-1998; 98GB-0006113.
XX      (BIOG-) BIOGENMA UK LTD.
XX      Wyatt P, Roberts JA, Whitelaw C;
PI      WPI: 1999-580449/49.
XX      P-PSDB: AAY42644.
DR      A nucleic acid encoding a signal transduction protein involved in plant
PT      dehiscence, useful for producing shatter resistant male sterile plants
PT      -
PS      Example 2; Fig 5; 71pp; English.
XX      The invention provides a nucleic acid encoding a signal transduction
XX      protein involved in the process of dehiscence. The nucleic acids and
XX      proteins are useful for regulating or controlling dehiscence of a pod or
XX      an anther in a plant, useful in the production of male sterile plants.
XX      The methods, etc. may be used in production of shatter resistance or
XX      shatter delayed plants such as oilseed rape (Brassica napus). The present
XX      sequence represents the promoter region and a partial coding region of
XX      B. napus DZ2B gene.
SQ      Sequence 1716 BP; 604 A; 256 C; 264 G; 576 T; 16 other;

Query Match          17.6%; Score 106.2; DB 20; Length 1716;
Best Local Similarity 83.9%; Pred. No. 1.1e-19;
Matches 120; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY      51 AATTAAGAAGAAACCTAAACGTTGATCGGAGATGATGATGATGATGATGATGATGATG 110
DB      1568 AAGTGAAGAGAAACTTAACGTTGATGATGATGATGATGATGATGATGATGATGATG 1627
QY      111 ATGAGAGATCATCAAAAGCATTGGGGTATTTCAAGACAGACGAAATTAAGGAGAGG 170
DB      1628 ACGAGAAATCATCAAAATGATCGGTGGAATTTCAAGACGCGTAAGAAAGGTGAGAGG 1687
QY      171 CAGTATCATCCACCGGTGACGCG 193
DB      1688 CAGTGAACATCCACGCGACGCG 1710

```

RESULT 9

```

AA79325
ID      AA79325 standard; cDNA; 1656 BP.
XX      AC      AA79325;
XX      27-NOV-2000 (first entry)
XX      Pinus radiata cell signalling involved polynucleotide seq ID NO:63.
XX      Eucalyptus grandis; Pinus radiata; Monterey pine; plant; modification;
XX      plant cell signalling; modulation; transgenic plant; pathogen; growth;
XX      environmental change; development; cell proliferation; differentiation;
XX      elongation; survival; disease resistance; nutrient metabolism; ss.
XX

```

OS Pinus radiata.
 XX
 PN WO200042171-A1.
 XX
 PD 20-JUL-2000.
 XX
 PF 11-JAN-2000; 2000WO-US00724.
 XX
 PR 12-JAN-1999; 99US-0228986.
 PR 01-NOV-1999; 99US-0162866.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PI Strabala TJ, Nieuwenhuizen NJ;
 XX
 DR WPI: 2000-476052/41.
 XX
 PT Isolated polynucleotide encoding a polypeptide involved in cell
 PT signaling used for generating transgenic plants with modified responses
 PT to external signals -
 XX
 PS Claim 1; Page 58; 527pp; English.
 XX
 CC AAA79263 to AAA79736 and AAB25100 to AAB25570 represent polynucleotide
 CC and protein sequences isolated from eucalyptus (Eucalyptus grandis) or
 CC pine (Pinus radiata also known as Monterey pine). The protein sequences
 CC are involved in cell signalling. The polynucleotide and protein
 CC sequences can be used to modify the response of plant cells to external
 CC signals e.g. environmental changes or pathogens during the growth and
 CC development of a plant. They can be used to modify cell proliferation,
 CC differentiation, elongation and survival, resistance to disease and
 CC nutrient metabolism. Examples of modifications which can be produced are
 CC altered fruit ripening and senescence of leaves and flowers e.g. to
 CC delay senescence and prolong the life of cut flowers or enhance
 CC senescence of reproductive organs to engineer sterile plants. Other
 CC modifications can be used to delay senescence in selected cell types or
 CC organs providing fruit and vegetables which have a longer shelf life
 CC between harvest and consumption, or to decrease branching frequency in
 CC forest tree species giving long stretches of valuable knot-free clear
 CC wood which can be used in solid timber furniture and veneers.
 CC
 SQ Sequence 1656 BP; 514 A; 283 C; 400 G; 459 T; 0 other;
 XX
 Query Match 9.6%; Score 58.2; DB 21; Length 1656;
 Best Local Similarity 55.6%; Pred. No. 2.6e-06;
 Matches 133; Conservative 0; Mismatches 103; Indels 3; Gaps 1;
 XX
 QY 72 TGTTCATGTCGATGATGATCCACTAACTTATATTCATGAGAGATCATCAAGCGA 131
 DB 551 TCTTCTCGTGGAGACACACAAATCAACAGATAATTTTCGGGAGGCTTCAAGCC 610
 QY 133 TTGGGGATATTTCACAGACGCGATTAACGGTGGAGGCGAGTAATCATCCACCGGAGC 191
 DB 611 TTAATCTTACTGTGAAGAACTGGAATGGGAAGTAGCAGTGAATTTTAAACGAG 670
 QY 192 GCGGCTCATCTTTGACCTTATCTAATGATAAAGAAATGCCGAGGAGGATGTGTTT 251
 DB 671 GCAG---AACATATGATCTTGTTAATGACAAAGAGATGCCGTATATGATGGGCAATG 727
 QY 252 CGCAACTAGAGCTAGAGAAATGAGATGAGTAAGTAAGTATGTTGGGTACTTCA 310
 DB 728 AGGCACAAAGCACTGATGATCAATGGAGTCAAGACACCCATTGTGACCTTACGCA 786
 XX
 RESULT 10
 AA222979
 ID AA222979 standard; DNA: 1324 BP.
 XX
 AC AA222979;
 XX
 DT 10-JAN-2000 (first entry)
 XX
 DE A. thaliana DZ2AT3 promoter region and partial coding sequence.

XX
 KM Signal transduction protein; deniscence; male sterile plant; DZ2AT3;
 KM Shatter resistance; oilseed rape; response regulator protein;
 KM Promoter; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN WO9949046-A1.
 XX
 PD 30-SEP-1999.
 XX
 PF 22-MAR-1999; 99WO-GE00905.
 XX
 PR 20-MAR-1998; 98GB-0006113.
 XX
 PA (BIOG-) BIOGEMA UK LTD.
 PI Wyatt P, Roberts JA, Whiteclaw C;
 XX
 DR WPI: 1999-580449/49.
 DR P-PSDB: AAY42648.
 XX
 PT A nucleic acid encoding a signal transduction protein involved in plant
 PT deniscence, useful for producing shatter resistant male sterile plants
 PT .
 XX
 PS Example 3; Fig 11; 71pp; English.
 XX
 CC The invention provides a nucleic acid encoding a signal transduction
 CC protein involved in the process of deniscence. The nucleic acids and
 CC proteins are useful for regulating or controlling deniscence of a pod or
 CC an anther in a plant, useful in the production of male sterile plants.
 CC The methods, etc. may be used in production of shatter resistance or
 CC shatter-delayed plants such as oilseed rape (Brassica napus). The present
 CC sequence represents the promoter region and a partial coding region of
 CC A. thaliana DZ2AT3 gene.
 CC
 SQ Sequence 1324 BP; 460 A; 214 C; 182 G; 467 T; 1 other;
 XX
 Query Match 8.1%; Score 48.8; DB 20; Length 1324;
 Best Local Similarity 71.3%; Pred. No. 0.001;
 Matches 87; Conservative 0; Mismatches 17; Indels 18; Gaps 1;
 XX
 QY 12 AATCGAANTGGCAACAAATCCATCGGAGATATCGGAAATAAAGAAAGAACTA----- 67
 DB 1192 AATCGAAGATGGCAACAAATAATCCACGAGATACCGGAAACAGTGGATGAAGTGA 1251
 QY 68 -----AACGTGTGATCGTCGATGATGATCCACATTAACCTTATATTCATG 113
 DB 1252 AGAAGAACTAATCAAGTGTGATCGATGATGATCCATTAACCGTAGCTCCAGC 1311
 QY 114 AG 115
 DB 1312 AG 1313
 XX
 RESULT 11
 AA73166/C
 ID AA73166 standard; DNA: 1965 BP.
 XX
 AC AA73166;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:27978.
 XX
 KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KM cytosolic; gene therapy; vaccine; metastasis; ds.
 OS Homo sapiens.
 XX
 PN WO200157182-A2.

PD 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US01354.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214986.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225265.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226272.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239936.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246615.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251860.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-483426/52.
DR

PE	17-JAN-2001:	2001WO-US01324.	XX					
PR	31-JAN-2000:	2000US-0179065.	PR	29-SEP-2000:	2000US-0236370.	PR		
PR	04-FEB-2000:	2000US-0180628.	PR	02-OCT-2000:	2000US-0236802.	PR		
PR	24-FEB-2000:	2000US-0184664.	PR	02-OCT-2000:	2000US-0237037.	PR		
PR	02-MAR-2000:	2000US-0186350.	PR	02-OCT-2000:	2000US-0237038.	PR		
PR	16-MAR-2000:	2000US-0189874.	PR	02-OCT-2000:	2000US-0237039.	PR		
PR	17-MAR-2000:	2000US-0190076.	PR	13-OCT-2000:	2000US-0237040.	PR		
PR	18-MAR-2000:	2000US-0198122.	PR	PR	13-OCT-2000:	2000US-0239935.	PR	
PR	19-MAR-2000:	2000US-0205125.	PR	PR	20-OCT-2000:	2000US-0239937.	PR	
PR	07-JUN-2000:	2000US-0209467.	PR	PR	20-OCT-2000:	2000US-0241221.	PR	
PR	28-JUN-2000:	2000US-0214886.	PR	PR	20-OCT-2000:	2000US-0241785.	PR	
PR	30-JUN-2000:	2000US-0215135.	PR	PR	20-OCT-2000:	2000US-0241786.	PR	
PR	07-JUL-2000:	2000US-0216647.	PR	PR	20-OCT-2000:	2000US-0241787.	PR	
PR	07-JUL-2000:	2000US-0216880.	PR	PR	20-OCT-2000:	2000US-0241809.	PR	
PR	11-JUL-2000:	2000US-0217487.	PR	PR	20-OCT-2000:	2000US-0241826.	PR	
PR	11-JUL-2000:	2000US-0217496.	PR	PR	PR	20-OCT-2000:	2000US-0244611.	PR
PR	14-JUL-2000:	2000US-0218290.	PR	PR	PR	08-NOV-2000:	2000US-0246474.	PR
PR	26-JUL-2000:	2000US-0220963.	PR	PR	PR	08-NOV-2000:	2000US-0246475.	PR
PR	26-JUL-2000:	2000US-0220964.	PR	PR	PR	08-NOV-2000:	2000US-0246476.	PR
PR	14-AUG-2000:	2000US-0224518.	PR	PR	PR	08-NOV-2000:	2000US-0246477.	PR
PR	14-AUG-2000:	2000US-0224519.	PR	PR	PR	08-NOV-2000:	2000US-0246478.	PR
PR	14-AUG-2000:	2000US-0225211.	PR	PR	PR	08-NOV-2000:	2000US-0246523.	PR
PR	14-AUG-2000:	2000US-0225214.	PR	PR	PR	08-NOV-2000:	2000US-0246524.	PR
PR	14-AUG-2000:	2000US-0225266.	PR	PR	PR	08-NOV-2000:	2000US-0246525.	PR
PR	14-AUG-2000:	2000US-0225267.	PR	PR	PR	08-NOV-2000:	2000US-0246526.	PR
PR	14-AUG-2000:	2000US-0225268.	PR	PR	PR	08-NOV-2000:	2000US-0246527.	PR
PR	14-AUG-2000:	2000US-0225270.	PR	PR	PR	08-NOV-2000:	2000US-0246528.	PR
PR	14-AUG-2000:	2000US-0225447.	PR	PR	PR	08-NOV-2000:	2000US-0246532.	PR
PR	14-AUG-2000:	2000US-0225757.	PR	PR	PR	08-NOV-2000:	2000US-0246609.	PR
PR	14-AUG-2000:	2000US-0225758.	PR	PR	PR	08-NOV-2000:	2000US-0246610.	PR
PR	14-AUG-2000:	2000US-0225759.	PR	PR	PR	08-NOV-2000:	2000US-0246611.	PR
PR	18-AUG-2000:	2000US-0226279.	PR	PR	PR	08-NOV-2000:	2000US-0246613.	PR
PR	22-AUG-2000:	2000US-0226681.	PR	PR	PR	17-NOV-2000:	2000US-0249207.	PR
PR	22-AUG-2000:	2000US-0226868.	PR	PR	PR	17-NOV-2000:	2000US-0249208.	PR
PR	22-AUG-2000:	2000US-0227182.	PR	PR	PR	17-NOV-2000:	2000US-0249209.	PR
PR	23-AUG-2000:	2000US-0227029.	PR	PR	PR	17-NOV-2000:	2000US-0249210.	PR
PR	30-AUG-2000:	2000US-0228924.	PR	PR	PR	17-NOV-2000:	2000US-0249211.	PR
PR	01-SEP-2000:	2000US-0229287.	PR	PR	PR	17-NOV-2000:	2000US-0249212.	PR
PR	01-SEP-2000:	2000US-0229343.	PR	PR	PR	17-NOV-2000:	2000US-0249213.	PR
PR	01-SEP-2000:	2000US-0229344.	PR	PR	PR	17-NOV-2000:	2000US-0249214.	PR
PR	01-SEP-2000:	2000US-0229345.	PR	PR	PR	17-NOV-2000:	2000US-0249215.	PR
PR	05-SEP-2000:	2000US-0229509.	PR	PR	PR	17-NOV-2000:	2000US-0249216.	PR
PR	05-SEP-2000:	2000US-0229513.	PR	PR	PR	17-NOV-2000:	2000US-0249217.	PR
PR	06-SEP-2000:	2						

PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases .
PS Disclosure: SEQ ID NO 2950; 986bp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 32206 BP; 9503 A; 5480 C; 7017 G; 10206 T; 0 other;
Query Match 7.4%; Score 44.8; DB 22; Length 32206;
Best Local Similarity 54.2%; Pred. No. 0.043;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 430 ATGCATATATATTTTATATATGGAACACACATATATACGCTAAGTGTATATGC 489
DB 22201 ATATATATATACATGATATATATATGATATATACATGATATATATATATGCT 22142
QY 490 ATAGATACCTGCATGCTGCTGTTTGAATTTAGGCTTTATCGTCCGATATATAA 549
DB 22141 ATATATACATGATATGATATATGATATATATACATGATATGATATATATAT 22082
QY 550 TCATGATAGTGTGCTTTAAGCTTAAATATTTAATAAGGCTT 597
DB 22081 ACATGCTATATGATGATATATATGATATATATATATATACGTGT 22034
RESULT 14
AAV83939
ID AAV83939 standard; DNA; 80595 BP.
XX
AC AAV83939;
XX
DT 03-MAR-1999 (first entry)
XX
DE HC-contig derived from normal human chromosome 10q25.2 region.
XX
KM Yeast artificial chromosome; YAC; probe: eukaryotic chromosome;
KM neocentromere; replication; extra-chromosomal element; segregation;
KM cell division; artificial chromosome; gene therapy; mardel1(10);
KM human artificial chromosome; transgenic; chromosome 10; 10q25.2; ss.
XX
OS Homo sapiens.
XX
PN WO9651790-A1.
XX
PD 19-NOV-1998.
XX
PF 13-MAY-1998; 98WO-AU00352.
XX
PR 26-AUG-1997; 97AU-0008791.
PR 13-MAY-1997; 97AU-0006784.
XX
PA (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
PI Cancellia MR, Choo K, Du Sart D;
XX
DR WPI: 1999-009773/01.
XX
PT New isolated nucleic acid comprising neocentromere sequences from
PT eukaryotic chromosome - used to produce replicable, segregating
PT artificial chromosomes that can carry large amounts of DNA for gene
therapy
XX
PS Claim 8; Fig 6; 540pp; English.
CC The present sequence represents the HC-contig derived from normal human
CC chromosome 10, 10q25.2 region. This region can be naturally mutated to

CC produce an unusual chromosomal marker referred to as mardel(10). The
CC mardel(10) marker is mitotically stable and contains a functional
CC neocentromere at a location regarded as non-centromeric. This
CC neocentromere maps to q25.2 on chromosome 10. The specification describes
CC nucleic acid sequences derived from a eukaryotic chromosome, including a
CC neocentromere or its functional derivative or hybrid, that are able, in
CC a competent cell, of replicating, acting as extra-chromosomal element
CC and segregating during cell division. The sequences can be used to
CC construct artificial chromosomes for use in gene therapy comprising a
CC replicable, segregating nucleic acid that confers a specific phenotype
CC on cells. Human artificial chromosomes can propagate in human cells and
CC carry large amounts of DNA (e.g. therapeutic genes), and, being
CC extra-chromosomal, they are not mutagenic. The artificial chromosomes
CC are also useful for generation of transgenic plants and animals, in
CC production of proteins and to make diagnostic reagents, e.g. for
CC expression of cytokines, receptors and growth factors, or to increase
CC the copy number of a gene in a cell. The constructs may also be
CC used for functional and structural analysis of chromosomes.
XX
SQ Sequence 80595 BP; 23183 A; 16613 C; 16824 G; 23975 T; 0 other;
Query Match 7.4%; Score 44.8; DB 20; Length 80595;
Best Local Similarity 54.2%; Pred. No. 0.06;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 430 ATGCATATATATTTTATATATGGAACACACATATATACGCTAAGTGTATATGC 489
DB 15588 ATACATTAATATGATATATATGATATATAGACATTAATATGATATATATATAC 15647
QY 490 ATAGATACCTGCATGCTGCTGTTTGAATTTAGGCTTTATCGTCCGATATATAA 549
DB 15648 ATAAATATGATATATATGATATATAGACATTAATATGATATATATATATAGACATTA 15707
QY 550 TCATGATAGTGTGCTTTAAGCTTAAATATTTAATAAGGCTT 597
DB 15708 ATATGATATATGCTATATATATAGACATTAATATGATATATATATATAT 15755
RESULT 15
AAK76447
ID AAK76447 standard; DNA; 4445 BP.
XX
AC AAK76447;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human Immune/haematopoietic antigen genomic sequence SEQ ID NO:31259.
XX
KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KM cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0189076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227109.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0228509.
PR 05-SEP-2000; 2000US-0228519.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0232063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259676.

(HUMA-) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -
XX
PS Disclosure: SEQ ID NO 31259; 3071bp + Sequence Listing; English.
XX
CC AAK5951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK62170 to AAK51921. (I) have cytosolic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 10:50:27 ; Search time 47 Seconds
(without alignments)
3161.878 Million cell updates/sec

Title: US-09-646-679-14
Perfect score: 605
Sequence: 1 ggcacgacgacgaatcgaaga.....aataaggttcctacc 605

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 38353 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Parents NA: *
1: /cgn2_6/pdata/1/ina/5A.COMB.seq: *
2: /cgn2_6/pdata/1/ina/6A.COMB.seq: *
3: /cgn2_6/pdata/1/ina/6A.COMB.seq: *
4: /cgn2_6/pdata/1/ina/6B.COMB.seq: *
5: /cgn2_6/pdata/1/ina/PCIOS.COMB.seq: *
6: /cgn2_6/pdata/1/ina/Backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	58.2	9.6	1656 4	US-09-228-986-63 Sequence 63, Appl
2	34	8.9	7218 1	US-08-232-463-14 Sequence 14, Appl
3	44.8	7.4	80595 4	US-09-078-294-3 Sequence 3, Appl
4	44	7.3	80246 4	US-09-078-294-4 Sequence 4, Appl
5	40	6.6	5014 4	US-09-381-862-3 Sequence 3, Appl
6	39.8	6.6	2551 1	US-08-486-013-70 Sequence 70, Appl
7	39.8	6.6	2551 2	US-08-482-279-70 Sequence 70, Appl
8	39.8	6.6	2551 3	US-08-342-268-70 Sequence 70, Appl
9	39.8	6.6	2551 4	US-09-015-968-70 Sequence 70, Appl
10	39.8	6.6	2551 5	US-09-397-386-70 Sequence 70, Appl
11	39.8	6.6	4586 3	US-09-031-563-26 Sequence 26, Appl
12	38.6	6.4	80246 4	US-09-078-294-4 Sequence 4, Appl
13	38.6	6.4	80595 4	US-09-078-294-3 Sequence 3, Appl
14	37.6	6.2	6727 3	US-08-629-643A-5 Sequence 5, Appl
15	37.6	6.2	6727 4	US-08-629-643A-5 Sequence 5, Appl
16	37.6	6.2	6727 5	US-08-629-643A-5 Sequence 5, Appl
17	35.4	5.9	1830 6	US-09-155-884-5 Sequence 5, Appl
18	34.6	5.7	2404 1	US-08-484-105-7 Sequence 7, Appl
19	34.6	5.7	2404 2	US-08-484-106-7 Sequence 7, Appl
20	34.6	5.7	2663 1	US-08-136-743B-3 Sequence 3, Appl
21	34.2	5.7	900 2	US-08-771-784-1 Sequence 1, Appl
22	34.2	5.7	900 3	US-09-076-756-1 Sequence 1, Appl
23	34.2	5.7	3334 4	US-09-255-984-1 Sequence 1, Appl
24	33.8	5.6	84495 4	US-09-797-906-3 Sequence 3, Appl
25	33.6	5.6	4718 3	US-08-936-135-9 Sequence 9, Appl
26	33.6	5.6	4733 3	US-08-936-135-11 Sequence 11, Appl
27	33.6	5.6	4769 3	US-08-936-135-13 Sequence 13, Appl

28	33.6	5.6	4784 3	US-08-936-135-15 Sequence 15, Appl
29	33.6	5.6	6152 4	US-08-973-462-1 Sequence 1, Appl
30	33.4	5.5	3098 1	US-08-447-500-1 Sequence 1, Appl
31	33.4	5.5	3098 1	US-08-454-087-1 Sequence 1, Appl
32	33.4	5.5	3098 1	US-08-447-408-1 Sequence 1, Appl
33	33.4	5.5	3098 1	US-08-453-866-1 Sequence 1, Appl
34	33.4	5.5	3098 3	US-08-185-359-1 Sequence 1, Appl
35	33.2	5.5	288 4	US-09-556-868-1 Sequence 1, Appl
36	33.2	5.5	2061 2	US-08-960-022-11 Sequence 11, Appl
37	32.6	5.4	870 2	US-08-420-629-9 Sequence 9, Appl
38	32.6	5.4	84495 4	US-09-797-906-3 Sequence 3, Appl
39	32.2	5.3	1255 4	US-09-227-357-52 Sequence 52, Appl
40	32.2	5.3	480 1	US-08-282-581-4 Sequence 4, Appl
41	32	5.3	480 1	US-08-550-544-4 Sequence 4, Appl
42	32	5.3	4084 3	US-08-866-340-1 Sequence 1, Appl
43	32	5.3	4460 4	US-09-103-875-4 Sequence 4, Appl
44	31.8	5.3	990 3	US-08-921-209-3 Sequence 3, Appl
45	31.8	5.3	990 3	US-09-411-763-3 Sequence 3, Appl

ALIGNMENTS

```

RESULT 1
US-09-228-986-63
Sequence 63, Application US/09228986
Patent No. 6359198
GENERAL INFORMATION:
APPLICANT: Strabala, Timothy
APPLICANT: Nieuwenhuizen, Niels
TITLE OF INVENTION: Compositions Isolated from Plant Cells
FILE REFERENCE: 11000/1020
CURRENT APPLICATION NUMBER: US/09/228,986
CURRENT FILING DATE: 1999-01-12
NUMBER OF SEQ ID NOS: 130
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 1656
TYPE: DNA
ORGANISM: Pinus radiata
US-09-228-986-63

Query Match          9.6% Score 58.2; DB 4; Length 1656;
Best Local Similarity 55.6%; Pred No. 1.1e-07;
Matches 133; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

QY 72 TGTGATCGTGCATGATGATCCATTAACCTTAATTCATGAGAATCATCAAGCGA 131
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 551 TCTTGCTCGTGAAGACACACAAATCAAGATTAATTCGAGGGTGTTCMAAGCC 610
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 132 TTGGGGGATTTGCACAGACGGAATACGAGGAGGAGGAGGATATCATCCACGCTGACG 191
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 611 TTATCTTACTTGTAAGAACTGAGAAATGGAAGTACGATGACTATTTCAAGCAGG 670
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 192 GCGGCTCATTTTGAACCTTAATCCATGATGAATGAAGATGCCAGAGGATGCTGTT 251
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 671 GCAG---AACATATGATCTTGTGTGTATGACAAAGAGATGCTGTATGATGGCATG 727
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 252 CGACACTAGAGAGCTTAAGAAATGGAAGTGAAGTCAATGATTTTGGGGGATTTCA 310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 728 AGGCACACAGGCACTGATGATCATGCGATGAGGATGACACCCATTTGCACACTACAGCA 786
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 2
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIDFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

```



```

NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22113-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match
Best Local Similarity 7.5%; Score 54; DB 1; Length 7218;
Matches 33; Conservative 220; Mismatches 185; Indels 0; Gaps 0;

QY 1 GGCACGACGAGATGAGATGGCAAAATCCATGGAGATATCCGAAATAAAGAA 60
DB 1425 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1366
QY 61 GAACTAAAGTGTGATCGTATGATGATCCACTAAACCTTAATTCATGAGAAGAT 120
DB 1365 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1306
QY 121 CATCAAGGATGGGGGATTTTCACACAGCCGATACGCTGAGAGGAGCAATCAT 180
DB 1305 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1246
QY 181 CCACCGTGAAGCGGCTCATCTTTGACCTTATCTTAATGATAAAGATCCCGAGAG 240
DB 1245 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1186
QY 241 GGATGGTGTGGACACTAAGACTAAGAAATGAGATGAGTCAATGATTTGGTGG 300
DB 1185 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1126
QY 301 GGTGACTGCTGCTGCAATGAGAGAGCGAGGAGCTTCTATGAGAGCTGACTTAA 360
DB 1125 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1066
QY 361 CCATGCTTGGCAAAACGCTTAAACAGACAGATCATCCCTGCTCAATAACCACTCAT 420
DB 1065 TCGCAAGCTCCCTGACCTGACGCAAGCTGGAATTAATCTGTGAGCGTATGGCAAC 1006
QY 421 GGATGCTGATGATATA 438
```

DB 1005 GAAGCAAAATAGTTATA 988

```

RESULT 3
US-09-078-294-3
Sequence 3, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 3
LENGTH: 80595
TYPE: DNA
ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match
Best Local Similarity 7.4%; Score 44.8; DB 4; Length 80595;
Matches 91; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 430 ATGATATATATTTATTTATTTGAAACACACATTAATAGCTGTAAGTGTATGTC 489
DB 1558 ATACATTAATATGATATATGATATGATATAGACATTAATATGATATATATAG 15647
QY 490 ATAGATACCTGCGATGCTGCTTTAGATTTAGGCTTCTTAACTCCGCTGATATATA 549
DB 15648 ATAAATATGATATATGCTATATAGACATTAATATGATATATGCTATATAGACATA 15707
QY 550 TCATGTAAGTGTGCTTAAAGCTTATTAATTTAAATAGGCTT 597
DB 15708 ATATGATATATGATATATAGACATTAATATGATATATGATATAT 15755
```

```

RESULT 4
US-09-078-294-4
Sequence 4, Application US/09078294
Patent No. 6265211
GENERAL INFORMATION:
APPLICANT: Choo, Kong-Hong Andy
APPLICANT: Du Sart, Desiree
APPLICANT: Cancilla, Michael R.
TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
FILE REFERENCE: Davies Col
CURRENT APPLICATION NUMBER: US/09/078,294
CURRENT FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 29
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 4
LENGTH: 80246
TYPE: DNA
ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match
Best Local Similarity 7.3%; Score 44; DB 4; Length 80246;
Matches 89; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 434 ATATATATTTATTTATTTAGCAAAACACATTAATAGCTTAAAGTGTATGATGATAG 493
DB 15605 ATAAATATGATATATGATATATGATATAGACATTAATATGATATATGATATAGACATA 15664
QY 494 ATATGCTGATGCTGCTGCTTTAGATTTAGGCTTCTTAACTCCGCTGATATATATCAT 553
DB 15665 ATATGATATATGCTGATATATGACATTAATATGATATATGCTATATAGACATTAATAT 15724
```


;; FILING DATE: 18-NOV-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/930,649
;; FILING DATE: 14-AUG-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/889,020
;; FILING DATE: 26-MAY-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Campbell, Cathryn A.
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: P-UC 3003
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 535-9001
;; TELEFAX: (619) 535-9949
;; INFORMATION FOR SEQ ID NO: 70:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2551 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-09-397-386-70

Query Match 6.6%; Score 39.8; DB 4; Length 2551;
Best Local Similarity 45.6%; Pred. No. 0.027; Indels 0; Gaps 0;
Matches 140; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 298 TGGGCTGACTCACTGGCTGCAATGAAGAGGCGCGGCTTTCATGGAAGCTGGACT 357
DB 3 TGAGACCAACTGCTGATATACGAAAGACATTTATGCTGTAATGAAGTTGGT 62
QY 358 TAACCATGCTTGGCAAAACCGTTAACCAAGGACACATATCCCTCTCATTAACCACT 417
DB 63 TCTGACTGATAGATCTTACCTACAGAGATTCAAGTGTGCTTGCCATTGAACAAT 122
QY 418 CATGATCGTATGATGATATATTTATTTATGTAACACACATATTAAGCTTAAGT 477
DB 123 AGTATATATGTTTAT 182
QY 478 GGTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537
DB 133 GGTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 242
QY 538 CGTAT 597
DB 243 GATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 302
QY 598 CCTGAC 604
DB 303 CACTTAC 309

RESULT 11
US-09-031-563-26
;; Sequence 26, Application US/09031563A
;; Patent No. 6022708
;; GENERAL INFORMATION:
;; APPLICANT: Frederic de Sauvage
;; APPLICANT: Arnon Rosenthal
;; TITLE OF INVENTION: Fused
;; FILE REFERENCE: P1272
;; CURRENT APPLICATION NUMBER: US/09/031,563A
;; CURRENT FILING DATE: 1998-02-26
;; NUMBER OF SEQ ID NOS: 27
;; SEQ ID NO 26
;; LENGTH: 4586
;; TYPE: DNA
;; ORGANISM: Drosophila virilis
US-09-031-563-26

Query Match 6.6%; Score 39.8; DB 3; Length 4586;
Best Local Similarity 52.0%; Pred. No. 0.034;
Matches 89; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 427 TTGATGATATATATATTTTATATATGGAACACATATTAACGCTTAAGTGTATGA 486
DB 4054 TTATGATGA 4113
QY 487 TGCATAGACTGCTGCTGCTTTTAAATTAAGGCTCTTTATCGCGTGATATA 546
DB 4114 TGTATGATTA 4173
QY 547 TAAATGATTAAGTTGCTTAAAGCTTAATTAATTAATTAATTAAGGTTT 597
DB 4174 TGTAT 4224

RESULT 12
US-09-078-294-4/C
;; Sequence 4, Application US/09078294
;; Patent No. 6265211
;; GENERAL INFORMATION:
;; APPLICANT: Choo, Kong-Hong Andy
;; APPLICANT: Du Sart, Desirée
;; APPLICANT: Cancilla, Michael R.
;; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;; FILE REFERENCE: Davies Col
;; CURRENT APPLICATION NUMBER: US/09/078,294
;; CURRENT FILING DATE: 1998-05-13
;; NUMBER OF SEQ ID NOS: 29
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 4
;; LENGTH: 80246
;; TYPE: DNA
;; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match 6.4%; Score 38.6; DB 4; Length 80246;
Best Local Similarity 52.9%; Pred. No. 0.23; Indels 0; Gaps 0;
Matches 83; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 434 ATATATATTTTATATATATGGAACACATATTAAGCTTAAATTAAGTATGATGATGAT 493
DB 15426 ATATATACAT 15367
QY 494 ATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 553
DB 15366 ATATATATTTATGCTTATATACATATATATATATATATATATATATATATAT 15307
QY 554 GTAAGTTGCTTAAAGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 590
DB 15306 GTGTATATGAT 15270

RESULT 13
US-09-078-294-3/C
;; Sequence 3, Application US/09078294
;; Patent No. 6265211
;; GENERAL INFORMATION:
;; APPLICANT: Choo, Kong-Hong Andy
;; APPLICANT: Du Sart, Desirée
;; APPLICANT: Cancilla, Michael R.
;; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
;; FILE REFERENCE: Davies Col
;; CURRENT APPLICATION NUMBER: US/09/078,294
;; CURRENT FILING DATE: 1998-05-13
;; NUMBER OF SEQ ID NOS: 29
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 3
;; LENGTH: 80595
;; TYPE: DNA
;; ORGANISM: Nucleotide sequence of HC-contig
US-09-078-294-3

Query Match 6.4%; Score 38.6; DB 4; Length 80595;
Best Local Similarity 52.9%; Pred. No. 0.23; Indels 0; Gaps 0;
Matches 83; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 7, 2002, 12:38:26 ; Search time 1692 seconds
(without alignments)
4826.039 Million cell updates/sec

Title: US-09-646-679-14

Perfect score: 605

Sequence: 1 ggcacagcagcagcagcaga.....aataagggttcctctacc 605

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	75.4	12.5	376	9	AM278862
2	63.2	10.4	568	10	BE822715
3	51.8	8.6	502	10	BE454895
4	50.2	8.3	732	10	BE602512
5	49.8	8.2	426	10	BM373897
6	47.6	7.9	666	10	BI509132
7	47.4	7.8	696	12	AQ488569
8	47.2	7.6	442	10	BM368320
9	45.8	7.5	1745	10	BM415113
10	45.4	7.5	1024	12	CNS030D0
11	45.4	7.5	1225	12	CNS0161D
12	45.2	7.5	525	12	AZ301735
13	45	7.4	541	12	AZ640886
14	44.8	7.4	610	12	AZ009832
15	44.6	7.4	853	10	BM516724
16	44.6	7.4	439	10	BM373028
17	44.6	7.4	946	12	CNS06M23

C 18	44.2	7.3	355	9	AM076928	AM076928 f103f10.Y
C 19	44.2	7.3	651	9	BS554734	BS554734 BS554734
C 20	44	7.3	474	12	BH105190	BH105190 RPT-24-3
C 21	43.8	7.2	576	12	AZ683492	AZ683492 RPT-24-1
C 22	43.8	7.2	610	12	AQ308813	AQ308813 CITHI-E1-
C 23	43.8	7.2	639	12	AZ600917	AZ600917 IMQ18M21
C 24	43.8	7.2	646	12	AQ308019	AQ308019 CITHI-E1-
C 25	43.6	7.2	443	12	B35082	B35082 HS-1026-B1-
C 26	43.6	7.2	405	12	AC0887556	AC0887556 HS-5557-A
C 27	43.6	7.2	536	10	BF000109	BF000109 7h17h03.x
C 28	43.6	7.2	712	12	AQ321139	AQ321139 RPT11-10
C 29	43.6	7.2	1085	12	CNS02EW4	AT028523 Tetradon
C 30	43.4	7.2	421	9	AW119129	AW119129 x08903.x
C 31	43.4	7.2	500	9	AU087616	AU087616 AU087616
C 32	43.4	7.2	538	12	B1936067	B1936067 PFEST0A2
C 33	43.4	7.2	563	12	AZ329701	AZ329701 IM0054D01
C 34	43.4	7.2	563	12	AZ515009	AZ515009 IM0054D01
C 35	43.2	7.1	857	9	AL514897	AL514897 AL514897
C 36	43.2	7.1	1101	12	CNS0039G	AL063921 Drosophila
C 37	43	7.1	508	12	AQ183850	AQ183850 HS-3209-A
C 38	43	7.1	533	12	BH105130	BH105130 RPT-24-4
C 39	42.8	7.1	412	12	AZ429160	AZ429160 IMQ212024
C 40	42.8	7.1	587	12	AQ985402	AQ985402 RPT-23-3
C 41	42.6	7.0	582	12	AC0588220	AC0588220 HS-5340-A
C 42	42.6	7.0	605	12	CNS02W2E	AL116527 Tetradon
C 43	42.6	7.0	660	12	AG055603	AG055603 Pan trogl
C 44	42.6	7.0	707	12	AZ369151	AZ369151 IM0119P01
C 45	42.6	7.0	773	12	CNS07210	AL426238 clone BAO

ALIGNMENTS

RESULT 1
AM278862
LOCUS
DEFINITION
sf99f11.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1019-4294.5' similar to TR:082445 082445 RESPONSE REGULATOR PROTEIN. ; mRNA sequence.

ACCESSION
AM278862
VERSION
AM278862.1 GI:6667411
KEYWORDS
SOURCE
ORGANISM
soybean.
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE
1 (bases 1 to 376)
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,D., Corvett,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Matlin,J., Beck,C., Wylie,T., Underwood,K., Stepec,M., Thelsting,B., Allen,M., Bowers,I., Peterson,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Rittler,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
Trace considered overall poor quality. This clone is available through: Resgen, Invitrogen Corp. 2130 South Memorial Parkway
Huntsville, AL 35891 For further information call: (800)-533-4363
or contact via email: ccs@resgen.com
Seq primer: -40RP from Gibco

FEATURES
source
1..376
/organism="Glycine max"


```

BASE COUNT
ORIGIN
133 a      66 C      88 g      89 T
      /db_xref="taxon:3847"
      /clone="GENOME SYSTEMS CLOVE ID: GM-cl019-4294"
      /clone_lib="Gm-cl019"
      /tissue_type="Immature seed coats of greenhouse grown
plants"
      /lab_host="DH10B (Gibco BRL)"
      /note="Vector: pSPORI1 (Life Technologies); Site_1: Not I;
Site_2: Sal I; This cDNA library was constructed from mRNA
isolated from Immature seed coats (200-300 mgs) of
greenhouse grown plants. The library was prepared using
the Life Technologies Superscript cDNA library
construction kit. Complementary DNA was synthesized from
mRNA using a poly (dT) sequence with a Not I restriction
site. Sal I linker adapters were ligated to the
blunt-ended cDNA fragments followed by Not I digestion.
The cDNA fragments were directionally cloned into the Not
I-Sal I restriction site of the pSPORI1 vector. The
ligated cDNA fragments were transformed into E.coli
ElectroMax DH10B host cells (Gibco BRL). This library was
constructed by Dr. Lila Vookin and Dr. Anu Khanna."

```

Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Medigian Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-wodkins@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or Info@genome systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(T/C/G)-3'.

Query Match	12.5%	Score 75.4	DB 9	Length 376
Best Local Similarity	54.6%	Pred. No. 9.5e-08		
Matches 199	Conservative 0	Mismatches 156	Indels 6	Gaps 2
QY	62	AAACTAAGCTGTGATCGTCATGATGATCCACTAAACCTATATATTCATGAGAAATC	121	
Db	26	AAATTAACACACCTGTAGTAGAGTCACATTAATAATCAACCGAAGATTCACAAACCTG	85	
QY	122	ATCAAAACGATTGGGGGTATTTCACAGACAGCGAATAACGCTGAGAGACGATATATC	161	
Db	86	TTGAGAGTGTGGATGAAATGAAAAATCAAGGAGGGAATATGCGCAAGAGCAGTGCACATT	145	
QY	182	CACCGTAGCGCGGCTCATCTTTTGACCTATATCCATATGATTAAGAAATCCCGAGAG	241	
Db	146	CATTGCCATGAC---AAAGATTGGACCTGATTCTCATGACATGATATGCCATATG	202	
QY	242	GATGCTTTTGCACAACTAAGAAAGCTAAGAGAAATGGAAGTGAAGTCATGATTGTTGG	301	
Db	203	AATGCAATTGAGCGAACAAAGGAACTTCGCTCAATGGGCAATGGAGCATGATTGTTGGT	262	
QY	302	GTGACTCATGCGCTGCACATTAAGAGGAGCGCCAGGCGCTTTCATGGAAGCTGACCTAAC	361	
Db	263	GTATCATCAACGCTGT---ACGGAAAGCGAATAATCAAAATTTATGGAAGCGGCGACTGAT	319	
QY	362	CATTGCTGGCAAAACGGTTAAACCAAGACGACCAAGATCATCCGTCATTAACCAACTG	418	
Db	320	GACTACATAGAGAAACCTTTGAACCAATTTTAAGCTTAGTCTACCTCTTGATTAAGATC	376	

BASE COUNT
ORIGIN

163 a 110 c 95 g 179 t 21 others

/note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 865 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Rezel, Center for Computational Genomics and Bioinformatics, University of Minnesota. <http://www.cbc.umn.edu/researchprojects/soybean/index.html>. Rereckling was performed by Genome Systems, St. Louis, <http://www.genomesystems.com>, and 3' sequencing by the Reck Center for Comparative and Functional Genomics, University of Illinois, <http://www.life.uiuc.edu/biotech/reck.html>. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

RESULT 2	568 bp	TRNA	linear	EST 24-MAY-2001
BE822715/c	BE822715	568 bp	TRNA	linear
LOCUS	BE822715	568 bp	TRNA	linear
DEFINITION	BE822715	568 bp	TRNA	linear
ACCESSION	BE822715	568 bp	TRNA	linear
VERSION	BE822715	568 bp	TRNA	linear
KEYWORDS	BE822715.1	568 bp	TRNA	linear
SOURCE	EST.	568 bp	TRNA	linear
ORGANISM	soybean.	568 bp	TRNA	linear
REFERENCE	Eukaryota, max.	568 bp	TRNA	linear
AUTHORS	Eukaryota, max.	568 bp	TRNA	linear
TITLE	Eukaryota, max.	568 bp	TRNA	linear
JOURNAL	Eukaryota, max.	568 bp	TRNA	linear
COMMENT	Eukaryota, max.	568 bp	TRNA	linear

QY 90 ATCCACAAACCTTTAAATCTATGAGAAATCATCAACAGATGTGGGGATTTACAA 149
Db 567 ATAAATACNNNNNAAGATTTCTCAAAAGCTGTGGAGAGTGTGGATGAAGAAAATCAAG 508
QY 150 CAGCGAATAACGCTGAGAGACAGTAATCATCAACCGTAGCGGGGCTATCTTTGACC 209
Db 507 GAGTNGAAATGGCCACAGACAGCAGGACATTC---TTGCCATNAGCAAGATTGTGACC 451
QY 210 TTATCTAATGATGAATAAGAAATGCCCGAGAGGATGTGTTGTGACACATTAAGAAGCTAA 269
Db 450 TGATTTCTCAATGAGCATAGATATGCCCATCATGATATGGCATTTAGGCGCAACAAAGAACTTC 391
QY 270 GAGCAATGCAAGTGAAGTCATGATGTGTGGGGTGAAGCTCAGTGCTGAACATGAAGAG 329
Db 390 GCTCATGGGCATTTGTAGCATGATTTGGTGTATCATCACCTGT---ACGGAAGCAG 334
QY 330 AGCGCAGGGCTTTATGAGAACTGACATTAACCATTTCTTGGGAAAACGGTTAACCAAG 389
Db 333 AATTAGCAAAATTTATGGAAGGGGACATGAGATGACTACATGAGAAAACCTTGAACATTT 274
QY 390 ACAAGATCATCCCTTCATTTAACCAACTCA 419


```
/clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
HYCDNA0009 (5 to 45 DAP)"
/tissue_type="5-45 DAP Spike"
/lab_host="SOLR"
```

Notes: Vector: lambdaZAP; Site.1: EcoRI; Site.2: XhoI; Plants were grown in the greenhouse at the University of California, Riverside (Fenton, St. Close, TJ Close). Whole spikes with awns trimmed were collected at 5, 10, 15, 20, 30 and 45 DAP (Fenton). Total RNA was prepared from each pool, equal quantities of all six RNA pools were combined, poly(A) RNA was purified from the mixture, one primary unamplified cDNA library was made, and 1 million pfu were *in vivo* excised to give plasmidscript SK(-) cDNA phagemids (Choi) in the TJ Close lab at the University of California, Riverside. Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch Atkins and Wink). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phed value 20 or above. For more details on library preparation and sequence analysis see

BASE COUNT	155 a	177 c	226 g	147 t	27 others
ORIGIN					

Query Match	8.3%	Score 50.2;	DB 10;	Length 732;
Best Local Similarity	49.3%;	Pred. No. 0.098;		
Matches 189;	Conservative	0;	Mismatches 188;	Indels 6;
				Gaps 2

QY 58 GAATAAACTAAAGCTGTGATCGTCGATCATATATCACTAAACCTTTAATCATATGAGA 117
 Db 33 GAGAAAGATCAAGGCTCTCTCTCATGTGAGAGACCAAGAGATCCACAGAGGTGTGGCGAGAGC 92
 QY 118 GATATCAAACGCATTTGGGGGATTTTCCACAGACACCGAATTAACGGTGAAGAGGCATTAAT 177
 Db 93 GGTCCTGATAGGGCCCGCGGTGTGAGAGCGAGACAGGCCCAAGAGCGCGCGAGGGCGGTGCG 152

27 178 CATCCACCCGACGGGGCTC---ATCTTTGACCTATCTTAATGGATTAAGAAATGCC 234
 Ddb 153 GCGGTGCGGACGCGCGCGCGGTATGATCTCATCTCAGCGACACGACGATGCC 212

OY 235 CGAGAGGGATGCTGTTCGACCACTAAGNAGCTAAGAAATGGAAGTCAATCAT 294
| | | | |
Dd 213 GGTCAATGAGCGCCACGAGCGACGAGCAGATCCGGCGATGSGGTGACCACGACCAT 272

Q7 295 TGTGGGGTACTCTACTGGCTGACATGAAGAGAGCGCCAGGCCTTCATGGAAAGCTGG 354
||| ||| | ||| | ||| ||| ||| |||
D6 273 CGTGCCGTTG---TTCAGCGAGACGCCCTCCGTCGGATGTCAGAGCCTTCATGCCGCCGG 329

OY 355 ACTTAACCATTCGTTGGCAAAAACCGTTAAACAAGGACATCATCCCTCATTAAACCA 414
| | | | | | | | | | | | | | | | | |
Db 330 AGCCGAGCACTTACGCCCCAAGCCGTGACCAAGAGAAGCTGGGCACACTTCTCTCCAA 389

QY 415 ACATCATGATCCTTGATGGATAT 437
 || | ||| | || |
 Db 390 GTTCGGGCTTGCCTAGCGGCAT 412

RESULT 5
BM373897
Office
DU372007

LOCUS	BM373897	426 bp	mRNA	linear	EST 10-JAN-2002
DEFINITION	Ebm03_3.SQ003.A24_R IGF Bartley Ebma03 library Hordeum vulgare cDNA				
ACCESSION	clone Ebma03-SQ003.A24 5', mRNA sequence.				
VERSION	BM373897				
KEYWORDS	BM373897.1	GI:18117287			
	EST.				

SOURCE	
ORGANISM	barley.
	Hordeum vulgare

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Hordeum.
1 (bases 1 to 426)
REFERENCE
1. *Journal of the Royal Microscopical Society*, 1962, 82, 1-10.

AUTHORS
 Realey, P., Liu, H., Caldwell, D., McCallum, N., Mule, S., Cardie, L.,
 Ramsey, L., Machary, G., Marshall, D.F.M. and Maugh, R.
 TITLE
 Development of Barley Transcriptome Resources
 JOURNAL
 Unpublished (2001)
 COMMENT
 Contact: Maugh R

Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731

Fax: 00 44 1382 562426
 Email: rwaughescri.sari.ac.uk
 All sequence has a phred quality score of 20 or over
 Seq primer: M13 reverse.
 Location/Qualifiers

```
/organism="Hordeum vulgare"  
/cultivar="Optic"  
/db_xref="taxon:4513"
```

```

/clone_lib="IGF Barley Ema03 library"
/tissue_type="Maternal tissue"
/dev_stage="8 days post anthesis"

```

/note="Vector: pSPORn1; site_1: Sal I; site_2: Not I; Non-normalised library, directionally cloned into pSPORn1. Derived from maternal tissue dissected from developing

plants. Developed as part of the barley transcriptome resources of BBSRC/SERAD funded cereal IGF (Investigating Gene Function) project."

ORIGIN	8.28	Score 49.8	DB 10	Length 426
Query Match				

Matches 124; Conservative 0; Mismatches 97; Indels 6; Gaps 1,
0y 194 GGCCTCATCTTTGACCTTATCCTAATGATAAGAAATGCCGAGAGGATGGTTCG 253

Db 3 GACGCGGCGTTGACATGTTTTGTTAGATAAGGACATGCCATCATGACTGGTCCCGAG 62

Oy 254 ACAACTAAGAGGCTAAGAGAAATGGAAGTGAATCATGATTGTTGGGGTGACTTCACTG 313

Db 63 GCAGTTGTGAACATCCGTGCATATGGGGGAAGGTGATGTGAACATTGTTGGGTTTTCGCC 122

Qy 314 GCTGACATGAAGAGGAGCGCAGGCGCTTTCATGGAAGCTGACCTTAACATTTGCTTGGCA 373

Db 123 GATACCATGCATGCA-----GGCGTTCATGAAIGCTGGTGCATGCTTTCCGTGCC 176

QY 374 AAACGTTAACCAAGACAAGATCATCCCTGCATTAACCACTCAT 420

Db 177 AAACCATGAGGAGGAGGTTCTCATACCTATCATTCAGGAGATCAT 223

BI509132	LOCUS	BI509132	266 bp	-mRNA	linear	EST 29-AUG-2001
DEFINITION	Bee Brain Normalized/Subtracted Library, BB17 Apis					

ACCESSION	BI509132
VERSION	BI509132.1
KEYWORDS	GI:15359506
EST.	
basepairs	
contigs	

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Apoidea; Apidae; Apis.

AUTHORS
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.
TITLE
Development of Barley Transcriptome Resources
JOURNAL
Unpublished (2001)
COMMENT
Contact: Maugh R
Unit of Genomics
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: rmaugh@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse

FEATURES
Source
Location/Qualifiers
1..442
/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBed01.S0002.N02"
/clone_lib="TGF Barley EBed01 library"
/tissue_type="Endosperm"
/dev_stage="6 days post anthesis"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from endosperm tissue dissected from developing
grains (6 days post anthesis) in glasshouse grown barley
plants. Developed as part of the barley transcriptome
resources of BBSRC/SEERAD funded cereal IGF (Investigating
Gene function) project."

BASE COUNT
ORIGIN
113 a 89 c 120 g 120 t

Query Match
Best Local Similarity 54.6%; Score 47.2; DB 10; Length 442;
Matches 119; Conservative 0; Mismatches 93; Indels 6; Gaps 1;

QY 203 TTGACCTTATCTATGATGAAGAAATGCCGAGAGGATGTGTTGACACATGAAG 262
|||||
DB 218 TTGTGATTTGTTTGTGATGAAGACATGCCCATCATGATGCTGCCGAGCATGTGG 277
QY 263 AAGCTAAGAGAAATGAGATGATGATGTTGGGTGACTTCACGCTGACAT 322
|||||
DB 278 AAGATCCCTGCTATGCGGGAAGTGAAGTGTGAGATTTTGGGGTCTGCGGATGATCAT 337
QY 323 GAAGAGAGAGCGAGCGCTTTCATGGAAGCTGACTTACCATGCTTGGCAAAACCGTTA 382
|||||
DB 338 GCAATGGA-----GGCATTCATGATGCTGCTGATGCTTTCGCGCCCAACCAATG 391
QY 383 ACCAAGGACAGATCATCCCTCATTAACCAACTCAT 420
|||||
DB 392 AGGAGGAGAGGTTCTCATCTATCATCAAGAGATCAT 429

RESULT 9
BM415113/c 1745 bp mRNA linear EST 28-JUN-2002
LOCUS
DEFINITION
OP20184 Mixed Stage EST's from Globodera pallida, the potato cyst
nematode Globodera pallida cDNA, mRNA sequence.
ACCESSION
BM415113
VERSION
BM415113.1 GI:18381463
KEYWORDS
EST.
SOURCE
Globodera pallida.
ORGANISM
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
Tylenchoidea; Heteroderidae; Heteroderinae; Globodera.
REFERENCE
1 (bases 1 to 1745)
Heer, J., Sosinski, B., Pokrzywa, R.M., Wary, A. and Oppeerman, C.
Mixed Stage EST's from Globodera pallida, the potato cyst nematode
Unpublished (2001)
JOURNAL
Contact: Oppeerman, C
Center for the Biology of Nematode Parasitism
NC State University, IACR-Rothamsted

FEATURES
Source
Location/Qualifiers
1..1745
/organism="Globodera pallida"
/db_xref="taxon:36090"
/clone_lib="Mixed Stage EST's from Globodera pallida, the
potato cyst nematode"
/note="Vector: lambda GT11; This is a collaborative effort
between IACR-Rothamsted and North Carolina State
University. The library was constructed from mixed stage
G. pallida in lambda GT11 by Paul Burroughs,
IACR-Rothamsted."

BASE COUNT
ORIGIN
439 a 472 c 262 g 547 t 25 others

Query Match
Best Local Similarity 55.3%; Score 45.8; DB 10; Length 1745;
Matches 89; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 430 ATGATATATATTTTATATATGGAACACACATTAACGCTATAGTGTATGATCC 489
|||||
DB 901 AT 842
QY 490 ATGATACCTGATGATGTGTGTTTGAATTAAGGCTCTTATCCCGGTATATATGA 549
|||||
DB 841 ATATATATATATATATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 782
QY 550 TCATGTAGCTGTGCTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACTTAACT 590
|||||
DB 781 ATATCTATATCTA 741

RESULT 10
CNS030D0/c 1024 bp DNA linear GSS 15-MAY-2000
LOCUS
DEFINITION
Tetradon nigroviridis genome survey sequence T7 end of clone
184002 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION
AL222093
VERSION
AL222093.1 GI:7880912
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetradon nigroviridis.
ORGANISM
Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphia; Acanthopterygii; Percormorpha; Tetradontiformes;
Tetradontidae; Tetradon.
1 (bases 1 to 1024)
Roest-Crollius, H., Jallion, O., Dasilva, C., Fizes, C., Fisher, C.,
Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater putterfish Tetradon nigroviridis
Unpublished
2 (bases 1 to 1024)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,
Bernot, A., Fizes, C., Wincker, P., Bottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1024)
Genoscope.

TITLE
JOURNAL
DEFINITION
AUTHORS
REFERENCE
1 (bases 1 to 1024)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,
Bernot, A., Fizes, C., Wincker, P., Bottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1024)
Genoscope.
TITLE
JOURNAL
DEFINITION
AUTHORS
REFERENCE
1 (bases 1 to 1024)
Roest-Crollius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C.,
Bernot, A., Fizes, C., Wincker, P., Bottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1024)
Genoscope.
COMMENT
Submitted (12-APR-2000) to the EMBL/Genbank/DBP databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.

[illegible]

BASE COUNT
ORIGIN

Query Match	7.5%	Score 45.2;	DB 12;	Length 525;
Best Local Similarity	50.0%	Pred. No. 1.5;		
Matches 89; Conservative	0;	Mismatches 89;	Indels 0;	Gaps 0

QY 420 TGGATGCTTGATGGATATAATTTTAAATTATATGAACACACATATAACTCTAAGTGT 479

Db 324 TGCGTCCTTAATATAATATAATATAATATAATATAATATAATATAATATAATATAATAT 383

QY 480 GATGATCATAGTACTTCATCGATGTGGTTTAGAATTAGSGTCTTATCCGCC 539
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 384 AATAATAATAATAATAATAATAATAATAATAATAATAATAATAATTNNNN 443

Dy 340 TGAATATAAATCAAGTGTGTCCTTAAGCATTAAATATTTAATAAGGTTT
||| ||| ||| ||| ||| ||| ||| ||| |||
Db 444 NNATATTNNNTATTATATTTTTTNNNTTTTATTTTATAATAATTTNNTTTT

RESULT 13	
AZ640886/c	
LOCUS	541 bp DNA linear GSS 14-DEC-2000
DEFINITION	M0503E1F Mouse 10kb plasmid UGCLM library Mus musculus genomic
ACCESSION	clone UGCLM0503E1 F, DNA sequence.

KEYWORDS	GSS.
SOURCE	house mouse
ORGANISM	Mus musculus

REFERENCE AUTHORS	Mammalia; Euterpia; 1 (bases 1 to 541) Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C.
----------------------	--

TITLE	Mouse whole genome scaffolding with paired end reads from 10kb
JOURNAL	Plasmid Inserts
COMMENT	Unpublished (2000)
	Contact: Robert B. Weiss

Rm. 309, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLU, UT
84112, USA
Tel.: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0503 row: E column: 11
Seq primer: CGTTGTAATAACGACGCGCAAT
Class: plasmid ends
High quality sequence stop: 541.

FEATURES
source

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGGCM0503E11"
/clone_libs="Mouse 10kb plasmid UGGCM library"
/sex="Male"
/lab_host="E. coli strain XL0-Gold, T1-resistant, F-"
/notes="Vector: pMD247; Purified genomic DNA from M.
musculus C57BL/6J male; from the Jackson

```

Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adsorbed DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD4 (g1147311414b1A12072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adsorbed mouse DNA was annealed to adsorbed vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT
ORIGIN

Query Match	7.4%	Score 45;	DB 12;	Length 541;
Best Local Similarity	48.8%;	Pred. No. 1.6;		
Matches 120; Conservative	0;	Mismatches 126;	Indels 0;	Gaps 0;

Dy
345 TGGAGCTGCACTTAACCATTCCTTGGCAAACCGTTAACCAAGCACAATCATCCCTC 404

Db
503 TGGAGTTAGTAGAATTNGCTTGCTTGCTTTGGCTCTTACAAGTAGCATTTTCAACTAGA 444

DQ 405 TCAATACCACTCATGGAGCGTTGAGGSATAAATATTATATTAATGGAACACACAGA 469
| | | | | | | | | | | | | |
DB 443 TGAAATGTGCATTAAATATATATATATATATATATATATATATATATATATATA 384

Db 383 TATATATGTATATATATATACACACTATATATGTAGTCTGTAATATATATTTA 324

Db 323 TATATGTCCTATATATATGCTATATATGTCCTGCTATATATATATATATATA 264

Db 263 TATATA 258

RESULT 14				
AZ009832/c				
LOCUS	AZ009832	610 bp	DNA	linear
				GSS 25-FEB-2000

ACCESSION	DNA sequence.
VERSION	AZ009832
NUMBERS	AZ009832.1 GI:7085216
	...

SOURCE ORGANISM	house mouse, <i>Mus musculus</i> ; Eukaryota; Mammalia; E
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

REFERENCE
1 (pages 1 to 610)
AUTHORS
Zhao, S., Niernan, W., Feldbiyum, T., Malek, J., Shatsman, S., Akınret
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.,
and Fraser, C. M.

TITLE	Mouse BAC End sequences from Library RPI-23
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPI-23-324H4.TV
	Contact: Shaying Zhao

Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong

